Essays on Local Economic Growth in India

Essays on Local Economic Growth in India

A dissertation presented

by

Paul Novosad

to

The Department of Political Economy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Political Economy and Government

Harvard University

Cambridge, Massachusetts

May 2013

Dissertation Advisor: *Author:*
Professor Asim Khwaja **Paul Novosad**

Essays on Local Economic Growth in India

Abstract

This dissertation explores how local circumstances affect local economic growth in India. The first chapter asks whether local politicians influence local economic growth. Using a regression discontinuity design, an economic census, and monthly stock prices, I find that economic outcomes are particularly poor in constituencies that were narrowly lost by the party in control of government resources. Using international survey data to classify industries by their dependence on government officials, I argue that this effect is driven by political control over the bureaucracy. The second chapter examines how natural resource wealth affects the local economy. Combining data on mineral deposits, global mineral prices and local economic activity, I am able to isolate the direct effect of natural resource wealth. In the cross-section, towns in resource rich areas are smaller, with larger mining sectors and smaller manufacturing and retail sectors. The causal time series results suggest that these effects may be due to unobserved aspects of natural advantage. Resource booms result in broad-based growth in towns up to 50km from the nearest mineral deposit. Rural areas are affected at a smaller radius, with growth in agroprocessing, but a decline in service sectors, suggesting a reallocation of both labor and government inputs toward mineral extraction. In the third chapter, I evaluate the economic impact of rural roads, focusing on a large scale road building program undertaken by the Indian government between 2000 and 2012. I take advantage of the allocation rules of the program to develop two methods to identify the causal effect of a rural road. I find that new paved roads lead to large increases in village employment. Roads lead to an increase in firm size, suggesting that firms are inefficiently small when transportation costs are high.

Contents

List of Tables

List of Figures

Acknowledgments

Thanks go out to my advisors Michael Kremer, Ed Glaeser and Alberto Alesina, and especially to the chair of my committee, Asim Khwaja. Needless to say, I benefited enormously from discussions with my classmates and many other researchers at Harvard University.

Mr. PC Mohanan of the Indian Ministry of Statistics has been invaluable in providing assistance with the Economic Census. I am grateful to Francesca Jensenius for sharing electoral data and to Sandesh Dhungana and Pranav Gupta for excellent research assistance. These projects received financial support from the Center for International Development and the Warburg Fund (Harvard University). The work on PMGSY was supported by National Science Foundation Doctoral Dissertation Improvement Grant Award Number 1156205.

I am very thankful for my family who supported me these six years and put up with increasingly improbable claims that I would be even more stressed if I had a real job. Finally, a big thank you to Sam Asher, my friend and collaborator on all this work, who arrived in the doldrums of my PhD, with the Economic Census in hand, and the promise of years of merging data and a light at the end of the tunnel. It was worth it, it was fun, and I could not have done this alone.

Chapter 1

Politics and Local Economic Growth[1]

1.1 Introduction

Does politics affect local economic outcomes? Political considerations affect the spatial distribution of government transfers (Albouy, 2009; Ansolabehere and Snyder, 2006) and public spending (Finan, 2004; Besley et al., 2004). Favored firms are more likely to receive credit from state banks (Cole, 2009; Khwaja and Mian, 2005; Carvalho, 2010) and corporate bailouts Faccio2006, and their valuations are tied to the fortunes of powerful politicians (Fisman, 2001).[2]

Whether these targeted effects add up to a substantive impact on local economies is not as well understood. Politically motivated government inputs could be important for local growth, or they could displace productive activity.[3] Alternately, they could have no local impact at all.[4]

[1] Co-authored with Sam Asher.

[2] Some additional examples include Aghion et al. (2009), Khemani (2007), Hoover and Pecorino (2005), Dixit and Londregan (1996) and Arulampalam et al. (2009).

[3] Studies of the effect of local government spending on private sector activity find both positive and negative estimates (Cohen et al., 2011; Ramey, 2011; Shoag, 2011). While politicians may be able to create local jobs (Carvalho, 2010), they may also disproportionately tax and embezzle funds from their supporters (Sukhtankar, 2012; Kasara, 2007).

[4] This is suggested by both Khwaja and Mian (2005) and Cole (2009), who find that politically motivated loans largely result in default.

We construct a state-constituency-level census of firms in India to measure the effect of one form of political favoritism on the local economy.[5] We examine the effect of having a political representative who is affiliated with the party in control of government.[6] State-level legislative constituencies provide a large number of elections with staggered years, within a common institutional framework. The empirical challenge is that locations that favor the dominant party may differ in many characteristics from locations that vote for opposition parties.[7] We use a regression discontinuity design to control for these differences. By comparing locations where aligned candidates narrowly won to those where aligned candidates narrowly lost, we focus our empirical test on places that are likely to be comparable on unobservables.

This paper makes four contributions. First, we show that political alignment has large effects on total employment. Second, we find no measurable effects on construction of public infrastructure or creation of government jobs. Third, we show that this employment is a net benefit to firms, as reflected by share prices. Finally, we find evidence that regulation is an important mechanism for these effects.

We find that private sector employment growth is 1.7 percentage points per year higher, over a seven year period, in locations with government-aligned politicians, relative to locations with non-aligned politicians.[8] We find no effect on employment in the public sector or state owned firms, nor do we find an effect on the construction of public infrastructure.

Employment growth is not a sufficient statistic for firm welfare; politicians could be forcing firms to make inefficient hiring decisions (Shleifer and Vishny, 1994). We examine

[5]Economic data is rarely reported at the level of political constituencies. Data collection tends to aim for representativeness at a higher level of aggregation, and the boundaries of legislative districts change frequently. Most economic data from India (NSS, ASI) are reported at the district level; districts on average contain 10 legislative constituencies.

[6]An analog in the U.S. is whether Republican congressional districts have better economic outcomes when Republicans are in control of Congress.

[7]For example, a caste-based party is likely to win support from areas dominated by lower castes. These areas will have a history of marginalization, and be on different growth paths.

[8]Alignment is a characteristic of the political representative of a location, rather than the location itself. Nevertheless for ease of exposition, we will also use the terms “aligned location” and “non-aligned location.”

stock returns around elections, to test whether firms receive higher valuations when their home constituencies become aligned. We find that stocks experience cumulative abnormal returns in the range of 12-15% in the month following state elections, when the local government-aligned candidate is victorious, suggesting that politician alignment is indeed good for local firms.

How do politicians affect the local economy? We consider three classes of mechanisms: (i) regulation; (ii) direct transfers; and (iii) supply of factors of production. We classify industries using international survey data, and test these mechanisms by exploring which industry classes are most affected by the alignment of local politicians. The effect of political alignment is largest in industries that depend most on bureaucratic inputs, indicated by use of licenses and permits, and frequent meetings with government officials. We do not find strong evidence supporting other channels. We find suggestive evidence that outcomes are particularly bad for constituencies narrowly lost by the governing party, suggesting that regulatory enforcement may be a more efficient tool at inhibiting economic activity than at promoting it.

Our results are consistent with a model of politicians who choose between policy levers to maximize future electoral outcomes, with consideration for the costs and returns of using those policy levers. The burdensome regulatory environment in India and political control of bureaucrats (Iyer and Mani, 2012; Pritchett, 2009) make it very cheap for politicians to control the enforcement of regulation at a local level. The ease of targeting and low cost of regulatory enforcement appear to make it a desirable strategic tool for Indian politicians.[9]

Section 1.2 provides the institutional context for our work, and a conceptual framework for thinking about the importance of political alignment. Section 1.3 describes the sources and construction of our data, and Section 1.4 explains the empirical strategy. Section 1.5 presents results, and section 1.6 discusses possible interpretations. Section 1.7 concludes.

[9] Other work shows that institutions can restrict the choice sets of politicians (Grembi et al., 2012; Ferraz and Monteiro, 2010; Wyplosz, 2012). In the Indian context, the absence of an effective institution of bureaucratic independence expands politicians' choice sets.

1.2 Background and conceptual framework

This section describes the electoral system in India and the mechanisms by which politicians can affect firms. A conceptual framework below specifies how we think about political decision-making, and why we would expect political alignment to be important.

1.2.1 State politics and firms in India

This paper focuses on state-level elections and local growth in Indian legislative constituencies, which are the political subdivisions of states. State governments are central actors in the allocation of government inputs.

The Indian constitution grants significant administrative and legislative power to state governments. States incur 57% of total expenditures, and have administrative control over police, provision of public goods, labor markets, land rights, money lending, state public services, and retail taxes. States operate their own civil services, and in practice state politicians exert a significant degree of control over federally-appointed bureaucrats assigned to their state (Iyer and Mani, 2012). Survey evidence suggests that among all levels of government, the majority of Indian citizens hold state governments responsible for provision of public goods and public safety (Chhibber et al., 2004).

State elections are characterized by a first-past-the-post system. Candidates compete in elections to represent single-member legislative constituencies; the candidate with a plurality wins the seat. The party with the largest number of seats in an election has the first opportunity to form a government; it may do so alone or as part of a coalition. If the party fails to form a majority, the party with the next highest number of seats may try to form a majority coalition. The characteristic of this system that we exploit is that the representative (Member of the Legislative Assembly, or MLA) of a given location may or may not be a member of the party in control of government. Indian elections between 1990 and 2005 were competitive. In addition to the two major national parties (Indian National Congress and Bharatiya Janata Party), several regional and caste-based parties experienced electoral success in state elections.

Firms in India have a high dependence on government and government-supplied inputs. Low quality or lack of public infrastructure is a major constraint on firms: in 2005, 36% of Indian firms reported that electricity supply was a major or severe obstacle to firm growth. In 1990, nearly all banks were operated by the state, making the government a monopolist supplier of formal credit.[10]

India's burdensome system of industrial regulation, known as the License Raj, required firms to get state approval before making any major production decisions, including expansion, entry, hiring and firing of workers and importing goods. While the 1990s were a period of significant liberalization, the regulatory burden on firms remained high by international standards through the study period (Panagariya, 2008). Finally, state-owned firms remain an important part of the economy.[11]

The evidence that Indian politicians use the policy levers at their disposal for personal and electoral gain is ample. Income support programs are captured by politicians (Besley et al., 2011; Jha et al., 2009). Cole (2009) finds large electoral cycles in the provision of agricultural credit. Khemani (2007) finds that fiscal transfers to states serve the electoral interests of the ruling party. Histories of the large business houses in India document the close relationship between important businessmen and politicians (Singh, 2011; McDonald, 2010).

1.2.2 Conceptual Framework

We think of the majority or governing party as making choices over the allocation of government inputs, with electoral goals in mind.[12] If a local candidate's electoral success is affected by the quality of government inputs in his constituency, the party has an incentive to

[10]By 2005 some privatization of banks had taken place but 54% of banking sector employment remained in state-owned banks.

[11]In 1990, the public sector and state-owned firms accounted for 18.8% of non-farm employment; by 2005 this number was 13.8%

[12]We focus on a two-party example here. Appendix A.2 describes how we extend our empirical strategy to take into account many parties and dynamic coalitions.

favor locations held by its own politicians, and to disfavor locations held by party opponents. As is supported by the literature cited above, we therefore expect aligned constituencies to receive more government inputs than non-aligned constituencies. Assuming some convex cost function of deviating from equal provision across all locations, we would expect this electoral strategy to be concentrated in swing constituencies. Locations narrowly won by the party should receive the most resources, and locations narrowly lost should receive the fewest. We generate these predictions from a simple model with rational voters, described in Appendix A.1.

Politically motivated government inputs do not necessarily lead to growth. A party will choose the policy levers that maximize electoral returns, which may or may not be those best for growth. Our empirical strategy tests whether economic outcomes are different between aligned and non-aligned constituencies. If we find an effect, we infer that (i) politicians have allocated government inputs differently across aligned and non-aligned locations; and (ii) those government inputs have an effect on economic outcomes.

We think of each policy lever as having a cost to the party, an electoral return, an economic effect, and a level of spatial targetability.[13] Political alignment should have the largest impact on policy levers that (i) are targetable at the constituency-level; (ii) are low cost; and (iii) have a high electoral return.[14] The tools empirically chosen by politicians are then informative of which policies have these characteristics.

Two papers find results closely related to this paper. Brollo and Nannicini (2012) find that municipalities with state-aligned incumbents receive greater transfers in election years than municipalities with non-aligned incumbents. They find evidence that this effect is driven by non-aligned municipalities, suggesting that the center is actively involved (by withholding

[13] For example, a general transfer program like a pension would be difficult to target to a single constituency. Procurement contracts can be targeted directly to firms, and thus are highly targetable. Road and electricity projects tend to affect multiple constituencies; they are targetable to the extent that geographically proximate constituencies have similar characteristics. In the case of swing constituencies, this is not likely to be true.

[14] The economic return is relevant to the extent that the party is concerned about welfare. We generally expect economic and electoral returns to be correlated, as voteres reward economic performance. However, highly salient government actions may have higher electoral returns than economic returns.

funds) even in jurisdictions where it does not hold formal power. Arulampalam et al. (2009) find that center to state transfers in India are higher when state parties are aligned with the federal coalition. We exploit a similar electoral logic, but focus on measuring the local economic impact of these political behaviors.

1.3 Data

This section describes how we constructed the economic census of firms, as well as the other data sources used in this paper.

The standard economic data sets used in India report data at the level of districts, which are approximately ten times larger than legislative constituencies. We matched the village and town-level Economic Census and Population Census of India to legislative constituencies, creating to our knowledge the first dataset linking economic and population outcomes to legislative elections.

The Indian Ministry of Statistics and Programme Implementation (MoSPI) conducted the 3rd, 4th and 5th Economic Censuses respectively in 1990, 1998 and 2005.[15] The Economic Census is a complete enumeration of all establishments except those engaged in crop production and plantation; there is no minimum firm size, and both formal and informal establishments are included.

The Census codes information on the location of the establishment (village for rural areas and ward-block for towns), the type of ownership, number of employees and some of their characteristics, source of electricity and finance, and the social group of the owner. The main product of the firm is also coded using the 4-digit National Industrial Classification (NIC). More detailed information on income or capital is not included. The main strengths of the data are its comprehensiveness, and rich detail on spatial location and industrial classification of firms.

We obtained location directories for the Economic Censuses, and then used a series of

[15] The 1st and 2nd were conducted in 1977 and 1980, while the 6th was planned to begin in 2012.

fuzzy matching algorithms to match villages and towns by name to the population censuses of 1991 and 2001.[16] We were able to match on average 2923 (62%) of towns and 515,114 (93%) of villages.

The NIC industrial classification system was updated in 2004; we created a manual correspondence between the NIC codes used in the different rounds of the Economic Census, which entailed grouping together categories which were not kept distinct in all the NIC classifications. We are left with 217 industry codes.

We use data from the Population Census of India in 1991 and 2001, which includes village and town demographic data, as well as information on local public goods (roads, electricity, schools and hospitals).

We obtained geographic coordinates for population census locations from ML Infomap and matched them to the bounding polygons of legislative constituencies. All population and economic census data were then aggregated to constituencies. We measure employment growth as change in constituency-level employment from 1990-98 and 1998-2005.

Election data for the period 1980-2007 were downloaded and cleaned from the web site of the Election Commission of India. We created a time series of political parties by manually matching party names, taking into account party fragmentation and consolidation. We constructed state coalition alliances, and poll and election dates from newspaper articles. The top panel of Figure 1.1 displays a map of the locations where we were able to match electoral data to economic data, and shows the split of aligned and non-aligned locations. The bottom panel identifies which of these locations had elections decided by margins within the optimal bandwidth used below.

As states follow distinct electoral calendars, we define electoral variables based on the first election in a state after the baseline measurement period.[17] We ignore additional

[16]The Economic Censuses of 1990 and 1998 were conducted with the house listing for the 1991 population census, while the 2005 Economic Census used codes from the 2001 population census.

[17]Example: For the Economic Census period 1990-98, we use elections from 1991 (Kerala, Tamil Nadu, WB), 1992 (Punjab), 1993 (Chhattisgarh, Himachal Pradesh, MP, Meghalaya, Mizoram, Nagaland, Rajasthan, Tripura), 1994 (AP, Karnataka) and 1995 (Arunachal, Bihar, Gujarat, Jharkhand, Maharashtra, Manipur, Orissa).

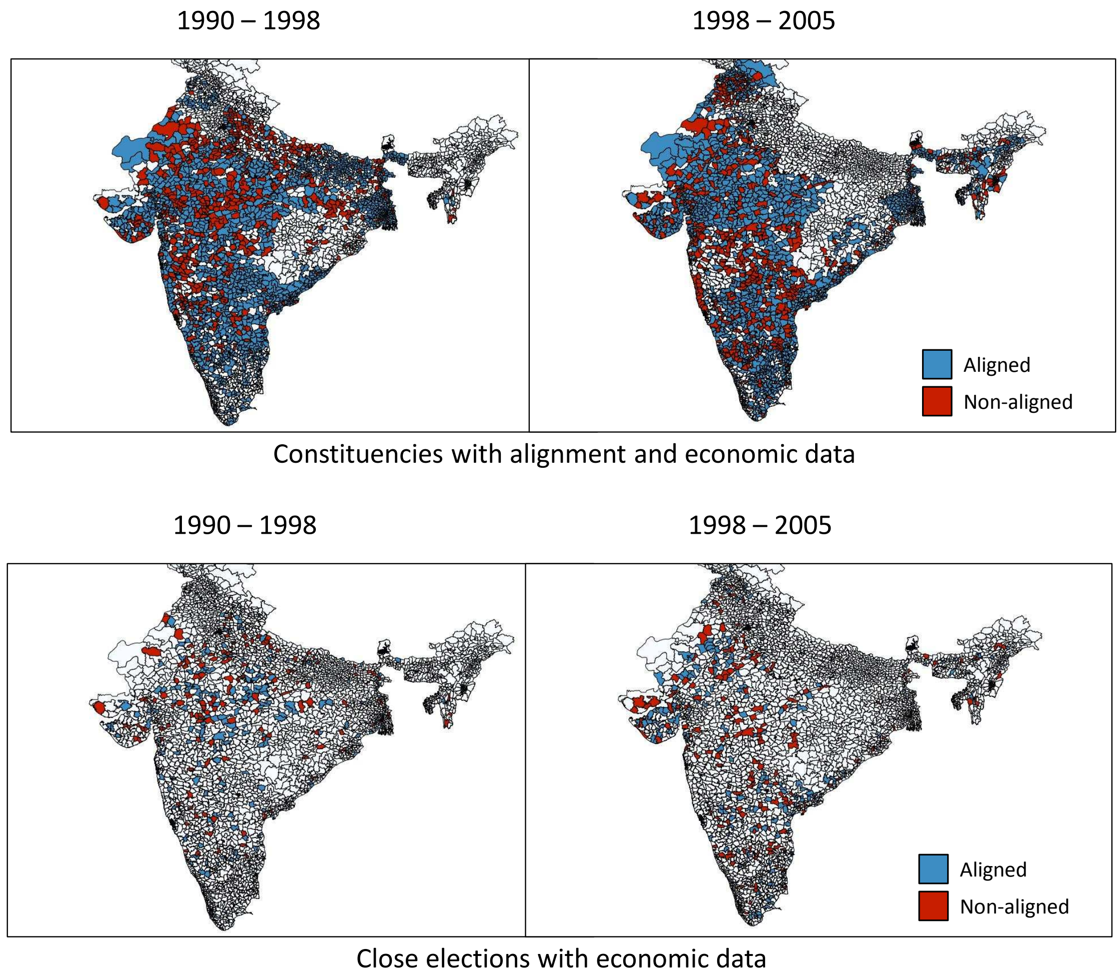


Figure 1.1: *Sample selection*

elections in the census period, and test robustness over different inclusion rules. Figure 1.2 illustrates this process. Incumbency conveys a zero or negative effect in Indian state politics (Uppal, 2009), so the exclusion of subsequent elections from the sample is not likely to create substantial bias through an incumbency channel. Given that the economic outcome periods span seven or eight years, our estimate of the effect of political alignment is biased downward to the extent that each observation includes several years without the identified politician being in power.

For stock prices and market indices, we use Datastream's monthly return index for

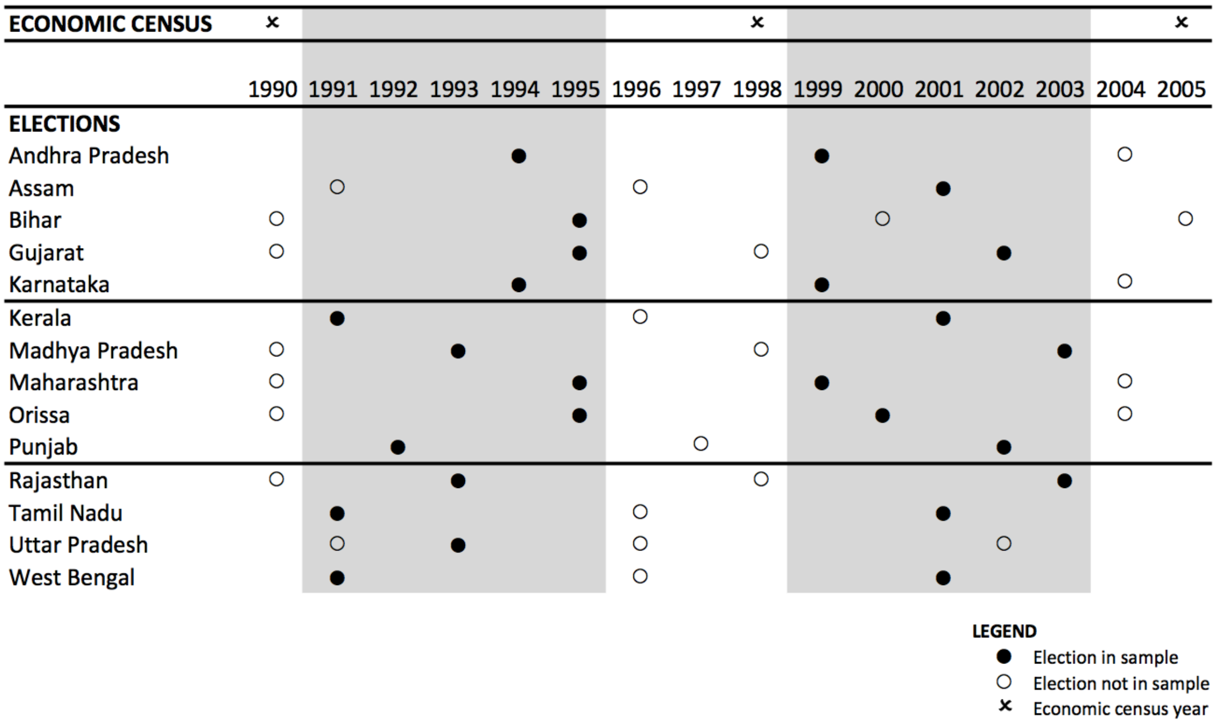


Figure 1.2: *Matching electoral variables to Economic Census rounds*

The figure shows the period of years used for construction of variables used from census and electoral data. The economic census was undertaken in 1990, 1998 and 2005. Elections happen at five-year intervals, with dates staggered across states. We explore changes in census values from 1990-98 and from 1998-2005. We match the first election in each state that occurred after the baseline observation period. We exclude elections in Uttar Pradesh in 1991 and 2002 because governments were very short lived. We exclude Assam 1991 because the dominant party was unregistered and ran as independents, making it impossible to code alignment. We exclude Bihar in 2000, because of the large number of post-electoral coalition changes.

individual equities listed on either the National Stock Exchange or the Bombay Stock Exchanges. We matched companies to legislative constituencies using the headquarter pincodes listed in Datastream, and pincode latitude and longitude from the GeoNames pincodes database. We limit the sample to companies located outside of India's major cities, as companies located in major cities are less likely to have a significant share of their operations in the constituency where their headquarters are located. We identified 166 firm-election pairs in 52 constituencies that experienced close elections between 1990 and 2007.[18]

[18]Before 1990, Indian stocks are very thinly traded. Our results are robust to alternate sample start dates. 2007

We construct industry-level measures of dependence on procurement, credit and bureaucratic inputs, using international data from the World Bank's Enterprise Surveys. These are based on firm-level surveys undertaken in 115 countries, including India, covering a range of topics about the business environment. We retain data on firms' revenue share from government, source of finance and quantity of external finance, and bureaucratic dependence, such as use of operating licenses and permits, and meetings with government officials, and paying of bribes. We construct measures both from the Indian survey and from the 2011 Standardized International Survey (which does not include India) to avoid any reverse causality from government policy to firm characteristics in India.

Our objective is to create an an industry-level indicator variable for each of these measures, that is equal to one if an industry has above the median value of all firms in that industry. In order to ensure that we are capturing cross-industry variation rather than cross-country variation, we first rank industries as high or low within a given country, and then take means of those measures across countries. We finally dichotomize these means to classify industries as high or low on each measure.

Table 1.1 shows constituency means of all variables at baseline, displayed separately for locations that end up with aligned and non-aligned politicians. The t statistic for the difference of means is displayed, as well as the t statistic from estimating Equation 1.1 with the baseline value as the dependent variable.

1.4 Empirical strategy

We would like to test whether locations with government-aligned politicians experience different economic outcomes from locations with non-aligned politicians.

---

is the terminal year, as electoral redistricting took place and we did not have access to nationwide geographic data on the new constituencies.

Table 1.1: *Summary Statistics*

Variable	Aligned constituencies	Non-aligned constituencies	t-stat on difference	RD estimate	t-stat on RD estimate
Baseline employment	14099	14292	0.22	-345	-0.21
Baseline public sector employment	2455	2436	-0.10	-34	-0.08
Baseline informal sector employment	12154	12302	0.15	342	0.26
Employment in firms of size <25	11546	11693	0.19	1277	1.21
Mean firm size	2.58	2.64	0.81	-0.22	-0.95
Number of establishments	5619	5750	0.33	588	1.28
Baseline population	172297	159042	-1.23	-515	-0.07
Number of villages	113	101	-1.13	3	0.30
Number of towns	1.05	1.08	0.39	0.09	0.37
Urban population share	0.48	0.47	-0.18	-0.02	-0.29
Share of villages with power supply	0.86	0.84	-0.58	0.06	1.72
Share of villages with tar access road	0.64	0.60	-0.74	-0.04	-1.33
Rural primary schools per village	0.87	0.87	0.04	-0.01	-0.52
Rural hospitals per village	0.03	0.03	-1.12	-0.00	-0.42
Share of land that is irrigated	0.15	0.22	1.74	0.02	0.53
Urban primary schools	14.11	14.22	0.09	5.16	1.46
Urban secondary schools	10.05	9.74	-0.33	2.06	1.10
Urban hospitals	1.38	1.34	-0.21	0.87	0.72
Urban electricity connections	5962	4931	-0.61	2043	1.23
Urban tar roads (km)	20	22	1.17	3	0.67

1.4.1 Local Economic Outcomes

We could run the following regression of constituency alignment on an economic outcome:

$$Y_{cst} = \beta_0 + \beta_1 * aligned_{cst} + \eta_s + \gamma_t + \epsilon_{cst}$$

Y_{cst} is an economic outcome in constituency c in state s at time t. $aligned_{cst}$ is an indicator for whether the politician representing constituency c is aligned with the governing party at the state level, and η_s and γ_t are state and year fixed effects. The term ϵ_{cst} refers to unobserved characteristics of the constituency or politician at time t.

The problem with this approach is that constituencies that elect politicians from the governing party may differ in unobserved ways from constituencies that elect opposition party politicians. At a minimum, the citizens of these constituencies have greater preferences for the dominant party, which could be correlated with many important economic factors.

To account for unobserved differences in politicians or constituencies, we focus on very close elections between aligned and non-aligned politicians. If these elections are close enough to swing for either candidate, they provide nearly random variation in the identity of the winning candidate (Lee, 2008; Lee and Lemieux, 2010). The underlying assumption of our regression discontinuity is that a constituency barely won by the aligned candidate is comparable to a constituency barely lost by the aligned candidate on any unobserved characteristics that could be correlated with the dependent variable. We run a standard set of tests of this assumption.

India is characterized by a large number of parties and candidates contesting elections. For simplicity, we first present our empirical design in a two-party context; further below we bring in the possibility of additional candidates and dynamic coalitions.

Consider a state with K constituencies, and candidates from two parties B and C contesting each electoral seat. The party which obtains a plurality of seats becomes the governing party.

In each constituency, let v^a represent the number of votes for the governing party, or aligned, candidate, v^n the votes for the non-aligned candidate, and v^{tot} the total number of votes. We define our running variable *margin* in constituency c at time t as

$$margin_{cst} = \frac{v^{a}_{cst} - v^{n}_{cst}}{v^{tot}_{cst}}.$$

Without loss of generality, let B be the governing party, so that candidates from party B are defined as aligned, and the definition of margin is:

$$margin_{cst} = \frac{v^{B}_{cst} - v^{C}_{cst}}{v^{tot}_{cst}}.$$

By construction, $margin_{cst}$ is positive if the candidate from party B has won the election in constituency c, and negative if B has lost. We can thus define the forcing variable $aligned_{cst}$ as an indicator equal to one if $margin_{cst}$ is greater than zero.

Since $margin_{cst}$ may covary with the outcome variable, we want to limit the test to locations with almost identical values of $margin_{cst}$. In the limit, these are constituencies where the election is decided fully at random. The population estimator β is defined by:

$$\beta = \lim_{m \to 0^{+}} \mathbb{E}[Y_i | margin_i = m] - \lim_{m \to 0^{-}} \mathbb{E}[Y_i | margin_i = m].$$

We use two standard specifications to generate sample estimates of this parameter, following Imbens and Lemieux (2008). Both tests estimate, separately for aligned and non-aligned constituencies, a regression of the outcome variable on *margin*. The predicted outcome is then compared across aligned and non-aligned constituencies.

The first test uses a local linear regression, with a bandwidth optimally calculated according to Imbens and Kalyanaraman (2009). We allow for the relationship between *margin* and the outcome variable to differ across aligned and non-aligned constituencies. The specification is described by Equation 1.1:

$$Y_{cst} = \beta_0 + \beta_1 align_{cst} + \beta_2 margin_{cst} + \beta_3 margin_{cst} * aligned_{cst} + \boldsymbol{\zeta} \mathbf{X}_{cst} + \gamma_t + \boldsymbol{\eta}_s + \epsilon_{cst}, \quad (1.1)$$

where Y_{cst} is a constituency-level economic outcome, $\mathbf{X}_{cst}$ is a vector of time-variant constituency controls, and $\boldsymbol{\eta}_s$ and $\mathbf{fl}_t$ are state and year fixed effects. ϵ_{cst} is clustered by state election. Constituency controls and fixed effects are not necessary for identification but improve the efficiency of the estimation. The effect of alignment with the governing party is identified by β_1.

The second test regresses the outcome variable on a polynomial function of the running variable *margin* across the entire sample of elections, and estimates a discontinuity at the point where *margin* becomes positive. The estimating equation is:

$$Y_{cst} = \beta_0 + \beta_1 * aligned_{cst} + f(margin_{cst}) + g(margin_{cst}) * aligned_{cst} + \boldsymbol{\zeta}\mathbf{X}_{cst} + \gamma_t + \boldsymbol{\eta}_s + \epsilon_{cst}, \tag{1.2}$$

where $f(\cdot)$ and $g(\cdot)$ are polynomial functions, and other variables are defined as in equation 1.1. The interaction between the polynomial function and $aligned_{cst}$ allows for a separate functional form for the running variable in aligned and non-aligned constituencies. β_1 estimates the effect of political alignment at the point where $margin_{cst} = 0$. ϵ_{cst} is clustered at the state-election level. Constituency-level controls included are log of total employment, log of population, urbanization rate and geographic size of constituency, with time variant controls measured in the baseline period.

State elections in India are contested by a large number of candidates and parties, and in more than half of cases the leading party needed to form a coalition in order to gain a majority. Appendix A.2 explains how we extend the empirical strategy above to account for more than two parties and dynamic coalition formation.

In brief, we assign parties to coalitions based on information known before the election takes place. We use newspaper articles or other documentation describing pre-election coalitions, or we predict coalitions based on alliance following the previous election if we could not find a description of pre-election alignment. This approach ensures that our result is not biased by the possibility that some unobserved factor (e.g. party competence) drives both entry into the coalition and the economic outcome. From this point forward, we use the term candidate alignment to mean predicted alignment rather than ex-post alignment.

We exclude constituencies where the top candidate ran as an independent, as we cannot observe whether independent candidates vote with or against the ruling coalition.[19]

1.4.2 Stock prices

Local employment growth alone does not imply that firms are better off: if firms' hiring decisions are a response to political pressure, employment growth could lower firm value. To directly test the effect of political alignment on firm value, we examine whether local politician alignment affects stock prices. In an efficient market, the prices of publicly traded firms reflect the information of all market participants. Firm prices can thus capture characteristics of firms that are unobservable to researchers, such as the value of political connections.[20]

If an election between an aligned and non-aligned candidate is expected to be close, the pre-election share price of a local firm will be a weighted mean of the value of the firm under an aligned politician, and the value of the firm under a non-aligned politician. After the election, the uncertainty is resolved and the share price reflects the value of the firm under the winning politician (Malatesta and Thompson, 1985). By comparing stock returns of firms in locations where the aligned candidate won with those where the aligned candidate lost, we estimate the value placed by the market on political alignment.[21]

If aligned politicians are forcing firms to hire workers when it is against the firms' interest, we should see stock prices fall when the local representative becomes aligned. However, if politicians are bringing useful government inputs to a location, we expect local stock prices to rise. We use a repeated "event study" methodology, using monthly

[19]Candidates from unofficial parties are reported by the Electoral Commission as independents, so cannot be distinguished from true independents and are excluded from the sample.

[20]This use of stock prices was first demonstrated by Roberts (1990) and has been used by Fisman (2001) and Jayachandran (2006). While the latter papers are based on direct relationships between politicians and individual firms (based respectively on family and political contributions), we focus on the inherent relationship between a firm's place of business and the local politician there.

[21]A single firm's price response to the election of an aligned candidate would capture a combination of the economic effect of alignment with the estimated ex ante probability of the aligned candidate winning. By comparing price movements of firms in locations that become aligned with those that become non-aligned, we will capture the full economic effect, as long as we have a sufficient number of firms in locations with close elections.

stock returns from India's two major stock exchanges, the Bombay Stock Exchange and the National Stock Exchange. We use monthly data because of the long lag between voting and official announcement of election results. Information is revealed throughout this period, so it is not possible to identify a single date when the information is processed by the market.[22]

For each event, we calculate cumulative abnormal returns as the residual from a market model estimated on a period from 24 months to 6 months prior to an election:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \nu_{it},$$

where R_{mt} is a value weighted market return index and ν_{it} is an orthogonal error term.

We estimate Equation 1.3 to determine whether politician alignment generates abnormal returns for local firms in the month following a close election:[23]

$$CAR_{i,t-1\rightarrow t+1} = \beta_0 + \beta_1 align_{i,t} + \beta_2 margin_{i,t} + \beta_3 margin_{i,t} * aligned_{i,t} + \boldsymbol{\zeta}\mathbf{X}_{i,t} + \boldsymbol{\eta}\mathbf{Y}_t + \boldsymbol{\gamma}\mathbf{S}_i + \epsilon_{i,t}, \tag{1.3}$$

where $CAR_{i,t-1\rightarrow t+1}$ is the cumulative abnormal return of a stock from the month before to the month following an election, and other variables are defined as in Equation 1.1. We limit the sample to close elections as above, cluster standard errors at the election level, and weight with a triangular kernel.

The inclusion of the *margin* variable in Equation 1.3 takes into account the fact that closer elections reveal more information to the market. If a winner was widely expected, we would expect an election to have no effect on asset prices. Estimating Equation 1.3 without the *margin* variable therefore puts a downward bias on β_1.[24]

[22]Voting often takes places on multiple days, and results may not be officially announced for days or weeks after voting ends. We define the end of our period as the last day of the month in which official electoral results were reported.

[23]Closeness of election in this case provides identification of the RD, and also implies that the local election result will be information that the market did not know before the election.

[24]Note that the win margin is an imperfect measure of the uncertainty over the result in advance of an election. For example, if an election turns out to be closer than expected, we are overestimating the ex ante closeness. However, we know of no data on advance polls or expectations of races for individual legislative constituencies, hence our use of win margin as a proxy for ex ante closeness.

1.4.3 Identifying the mechanisms

Our final objective is to identify what types of firms are most affected by political alignment. Our basic approach is to estimate Equation 1.1, with location or industry characteristics interacted with the electoral results. The factors that strengthen the effect of political alignment are informative about the constraints on firm growth, as well as on the policy tools being used by politicians. As most of the interaction measures we use are ordinal, we divide our firm or location sample into two groups, and code the ordinal characteristic with a dummy variable.[25]

Using procurement as an example, the estimating equation is:

$$\begin{aligned} Y_{csti} = \beta_0 + \beta_1 align_{cst} + \beta_2 margin_{cst} + \beta_3 margin_{cst} * aligned_{cst} + \beta_4 PROC_i + \\ \beta_5 align_{cst} * PROC_i + \beta_6 margin_{cst} * PROC_i + \beta_7 margin_{cst} * aligned_{cst} * PROC_i + \\ \boldsymbol{\zeta} \mathbf{X}_{csti} + \gamma_t + \boldsymbol{\eta}_s + \epsilon_{csti}, \end{aligned} \tag{1.4}$$

where $PROC$ is an indicator for whether a firm is in an industry with above average dependence on government procurement contracts, and i indexes the high and low procurement groups of firms. The other interactions used are proxies for dependence on bureaucratic inputs and dependence on external finance.

Note that the industry or location attribute that we interact with the dependent variable is not randomly assigned. Therefore it is possible that any interaction effect that we pick up is in fact driven by some other unobserved source of variation that is correlated with the interaction variable.

[25] An alternate approach would be to create a constituency-industry panel using the 217 product classification codes provided in the Economic Census. While this methodology is also sound, it produces extremely noisy estimates due to (i) the large number of zeroes in the full constituency-industry panel, since most constituencies do not produce most products; (ii) the large variation in the size of industries within a given constituency; and (iii) narrow misclassification of certain industries from one census period to the next. For example, coding appears to be inconsistent across the categories of "Manufacture of made-up textile articles (except apparel)" and "Manufacture of blankets, shawls, carpets, rugs and other similar goods." The result is that some locations show large growth in one industry and large decline in the other, when it is more likely only the classification has changed. However, both of these are classified as industries with low dependence on procurement, so much of this noise is eliminated when we aggregate to constituency-industry groups.

1.4.4 Balance Tests

The regression discontinuity approach relies on constituencies barely won by majority aligned candidates being alike on unobservables to constituencies barely lost by aligned candidates. This notion is challenged by recent work by Grimmer et al. (2012), who find that candidates who enjoy structural advantages in U.S. elections disproportionately win elections that are very close.[26] In light of this evidence, we take extra care to perform a large number of tests to demonstrate that these types of advantages do not drive the outcomes of close elections in India.

A key assumption for identification in the regression discontinuity model is that the forcing variable is continuous near the treatment threshold (Lee, 2008). Figure 1.3 describes the density of the forcing variable, *margin*. Constituencies with $margin > 0$ are those that were narrowly won by governing party-aligned MLAs, while those with $margin < 0$ were narrowly lost by aligned MLAs. Panel A shows the distribution of the win margin across our entire sample of Indian elections from 1980-2003. Panel B restricts the range to races with win margins of less than 5% and shrinks the bin size to focus on the discontinuity in the forcing variable. Both graphs suggest the density is continuous at zero. The mode of the *margin* distribution is to the right of zero because on average the ruling coalition wins more often than it loses.

Panel C shows the fit of a McCrary test of continuity in the density of the running variable at zero, for our full sample of elections (McCrary, 2008). Panel D restricts the test to the sample of elections matched to the population and economic census and used in subsequent analysis. The tests do not reject continuity in the running variable at the alignment threshold.

Figures 1.4 and 1.5 perform tests analogous to those performed by Grimmer et al. (2012). We analyze the tendency of close elections to be won or lost by candidates with two types of structural advantage: local incumbency, or membership in the state majority party, that is,

[26] Examples of structural advantages include alignment with the state majority party, the state Governor, or the Secretary of State's office.

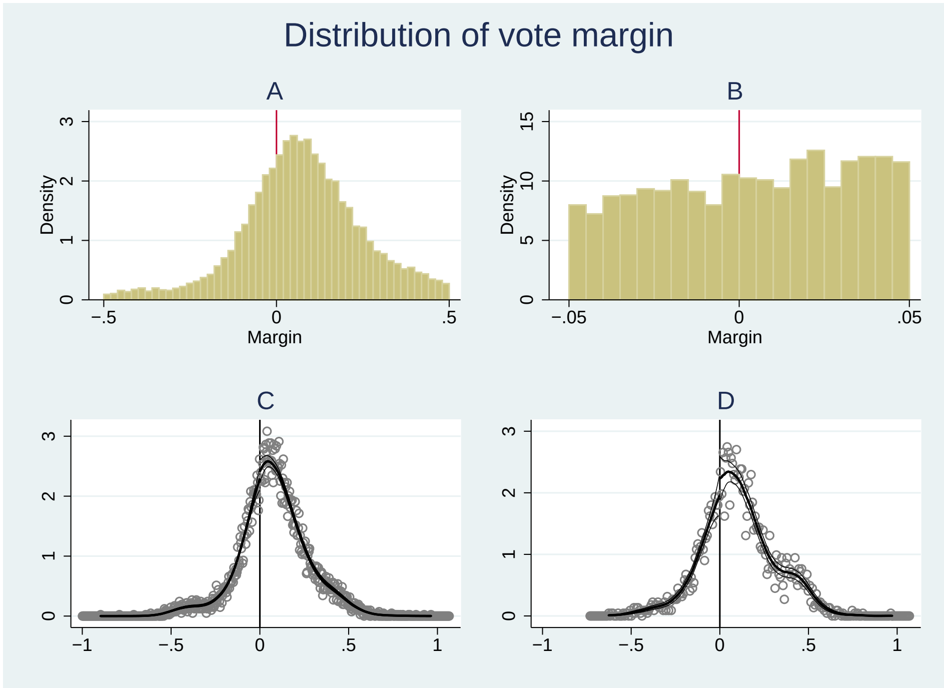


Figure 1.3: *Continuity of forcing variable*

The figure shows the distribution of electoral win margin, defined as vote share of the top candidate aligned with the coalition ruling party minus the vote share of the top non-aligned candidate. Panel A is a histogram of this margin across all Indian elections from 1980-2003. Panel B takes the same source data but focuses on the range of margins between -5% and +5%. Panel C plots a non-parametric regression to the left- and right-hand sides of the same data following (McCrary, 2008), testing for a discontinuity at zero. Panel D restricts the McCrary test to the sample used in analysis, which are the locations we were able to match to the Economic Census.

the party in control of government institutions when the election takes place. Figure 1.4 shows the McCrary tests of the win/loss margins of these two types of candidates; the tests do not reject continuity at the winning threshold. Figure 1.5 shows that the winners of very close elections are not more likely to have structural advantages than the losers. In fact, the left panel suggests that the winners of close elections are slightly less likely to be local incumbents.

If a close election provides variation in the winning candidate that is as good as random assignment, then constituencies narrowly lost by aligned candidates should be indistinguishable on observables from constituencies narrowly won by aligned candidates. Columns 4 and 5 of Table 1.1 shows the point estimate and t statistics from estimating equation 1.1

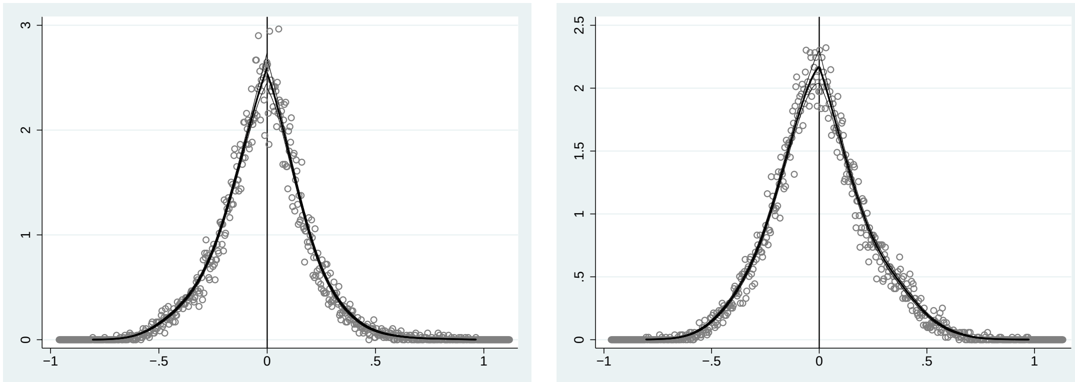


Figure 1.4: *Density functions of electoral performance of potentially structurally advantaged candidates*

The figure shows the density function of the margin of victory of candidates with a potential structural advantage in an election, as defined in Grimmer et al. (2012), along with a non-parametric smoother and its 95% confidence interval. In the left panel, the X axis is defined as the vote share of the local incumbent minus the top ranking non-incumbent. In the right panel, the X axis is defined as the vote share of the local majority-aligned candidate (from the previous election) minus the top ranking non-incumbent. The sample is all candidates who ran for a legislative seat between 1990 and 2005. If incumbents or majority parties enjoy a structural advantage in close elections, we would expect to see a discontinuity around zero.

on baseline constituency characteristics. The coefficient on the forcing variable *aligned* is significant at the 10% level in only one of these cases (rural electrification), indicating that aligned and non-aligned constituencies are alike on observables.

1.5 Results

This section provides evidence that political alignment has a significant positive effect on local private sector employment growth, and stock prices of firms. These findings are robust to a range of regression discontinuity specifications. This effect is largest in firms with a high dependence on bureaucratic inputs.

1.5.1 Economic outcomes

Table 1.2 presents the regression discontinuity estimates of the effect of political alignment on constituency-level log employment growth. Column 1 presents local linear regression estimates from Equation 1.1 with year and state fixed effects. The estimate on *aligned*

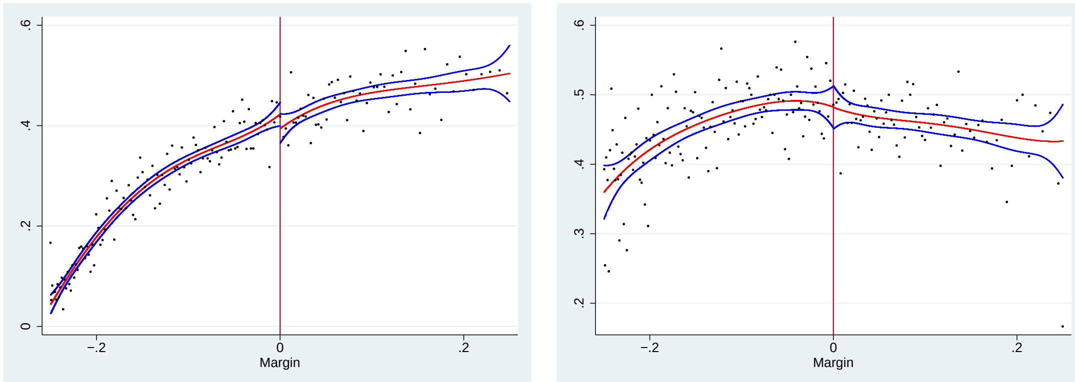


Figure 1.5: *Mean structural advantage of candidates and margin of victory/loss*

The figure plots the share of candidates with a potential electoral structural advantage (as defined in Grimmer et al. (2012), against their margin of victory or loss. In both panels, margin of victory is defined as the vote share of the local winner minus the 2nd place candidate. Within each percentage point sized bin, the point indicates the share of candidates with that result who were (left panel) incumbents or (right panel) members of the majority party. The sample is all candidates who ran for a legislative seat between 1990 and 2005. If these types of candidates enjoy structural advantages in close elections, these shares should be discontinuously higher to the right of zero.

indicates that where elections were closest, constituencies with aligned MLAs grew 1.6 log points more per year than non-aligned constituencies, over a seven year period. The measured effect over the 7- or 8-year census period is 12 percentage points.

Controls are not necessary for identification, but their inclusion increases the efficiency of the estimator. Column 2 adds lagged constituency controls. The annualized estimate falls to 1.5 log points, with a smaller standard error. To generate a measure indicative of total employment in the sample, column 3 weights observations by lagged employment. The estimate and standard error are unchanged, suggesting that the effect of political alignment is unaffected by constituency size.

Columns 4 through 6 present analogous estimates of the full sample polynomial specification (Equation 1.2) on the same outcome. The polynomial specification reports smaller estimates close to 1.0 percentage points per year, with similar statistical significance to the local linear specification.[27]

[27] Appendix Table A.1 runs standard placebo tests of these regressions, with simulated discontinuities at the 1st and 3rd quartile of the distribution of the win margin. As expected, these estimates are insignificant and close to zero. Appendix Table A.2 uses local linear regression with polynomial controls for the win margin on

Table 1.2: *Effect of politician alignment on log employment growth*

	Local linear regression			Polynomial regression		
	(1)	(2)	(3)	(4)	(5)	(6)
Aligned (RD)	0.016	0.015	0.015	0.011	0.010	0.009
	(0.008)*	(0.007)**	(0.007)**	(0.005)**	(0.005)**	(0.005)*
Margin	-0.248	-0.239	-0.260	-0.157	-0.150	-0.144
	(0.234)	(0.230)	(0.216)	(0.049)***	(0.044)***	(0.043)***
Margin * Aligned	0.195	0.117	0.139	0.134	0.134	0.132
	(0.266)	(0.269)	(0.252)	(0.039)***	(0.050)**	(0.047)***
Baseline		-0.023	-0.023		-0.018	-0.018
		(0.005)***	(0.005)***		(0.003)***	(0.003)***
Constant	0.000	0.202	0.197	0.054	0.193	0.194
	(0.010)	(0.044)***	(0.043)***	(0.002)***	(0.027)***	(0.027)***
Weighted	No	No	Yes	No	No	Yes
Controls	No	Yes	Yes	No	Yes	Yes
N	663	663	663	3625	3625	3625
r2	0.13	0.26	0.25	0.09	0.20	0.19

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. The table shows kernel regression discontinuity estimates of the effect of politician alignment with the state-level governing party on annualized constituency log employment growth from 1990-98 and 1998-2005. Columns 1-3 present local linear estimates (Equation 1.1), and columns 4-6 present full sample polynomial estimates (Equation 1.2). Column 1 is the baseline regression on year and state fixed effects. Column 2 adds lagged constituency controls, and column 3 weights observations by baseline employment. Columns 4-6 follow the same pattern. Standard errors are clustered at the state-election level.

Figure 1.6 plots estimates from Equation 1.1 with a range of bandwidths, an alternate kernel and a different window of election years. The local linear results are very stable. The full polynomial estimates in Panel D are affected by the inclusion or exclusion of observations with elections that were not close at all. The point estimate ranges from 0.011 with the inclusion of win margins smaller than 60%, and rises monotonically to 0.021 as the sample is restricted to smaller win margins. The premise of the regression discontinuity approach is that the identified treatment effect is only valid for very close elections; this sensitivity to inclusion of observations with very large win margins makes the polynomial specification less desirable, motivating us to focus on the local linear method for the analysis

both sides of the threshold. Estimates are very similar, albeit more noisy when a local 3rd degree polynomial is used.

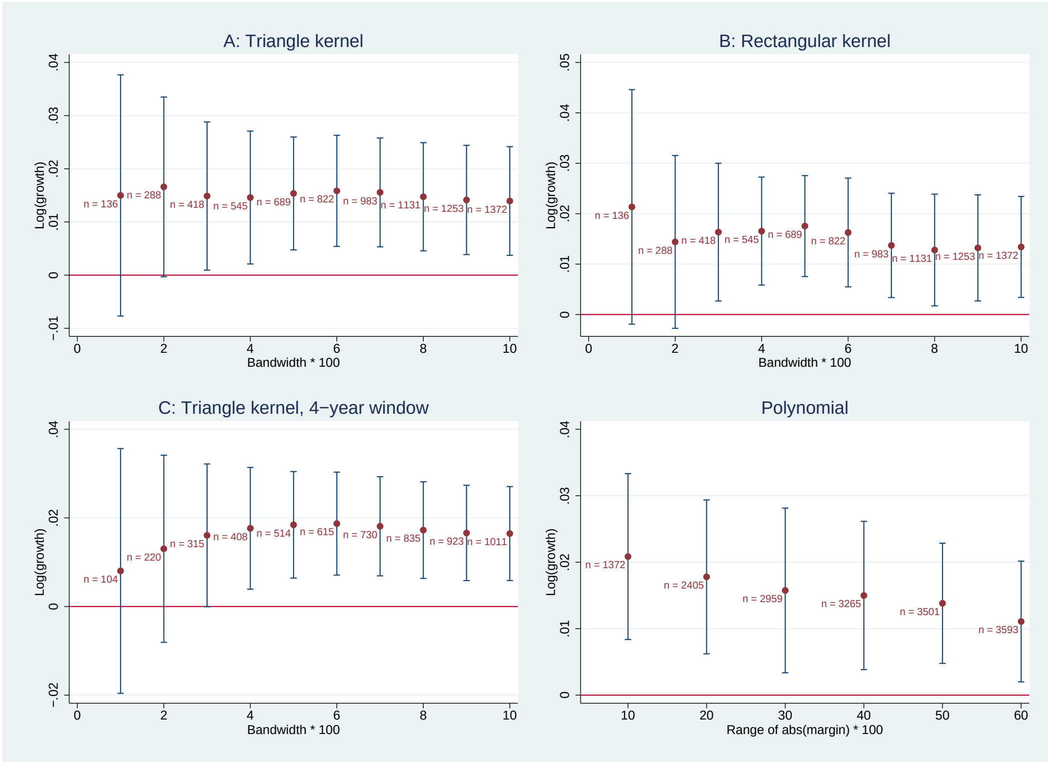


Figure 1.6: *Robustness of employment effect to alternate specifications*

This figure plots the point estimate and 95% confidence interval of the constituency-level treatment effect of alignment with governing party on annualized log employment growth under a range of specifications. Panel A shows the treatment effect of equation 1.1 with bandwidth on the x-axis. Panel B repeats Panel A using a rectangular kernel. Panel C repeats Panel A, but using a 4-year window of elections instead of a 5-year window of elections. Panel D shows the effect of limiting the range of the running variable when running the polynomial specification in equation 1.2. In each case, the x-axis shows in percentage points the vote share of the aligned candidate minus the vote share of the non-aligned candidate.

below. Reassuringly, the polynomial estimates are positive and highly significant for all samples used, and the local linear estimates fall at the midpoint of the range of polynomial estimates.

Figure 1.7 presents a visual representation of the regression discontinuity estimates. The win margin for the aligned candidate is plotted on the x-axis, with log employment growth on the y-axis. The sample is divided into bins based on the win margin of the aligned candidate, and the points represent mean log employment growth in those bins. Locations just to the right of the solid vertical line were narrowly won by aligned candidates, while locations just to the left of the solid line were narrowly lost. The regression lines show the

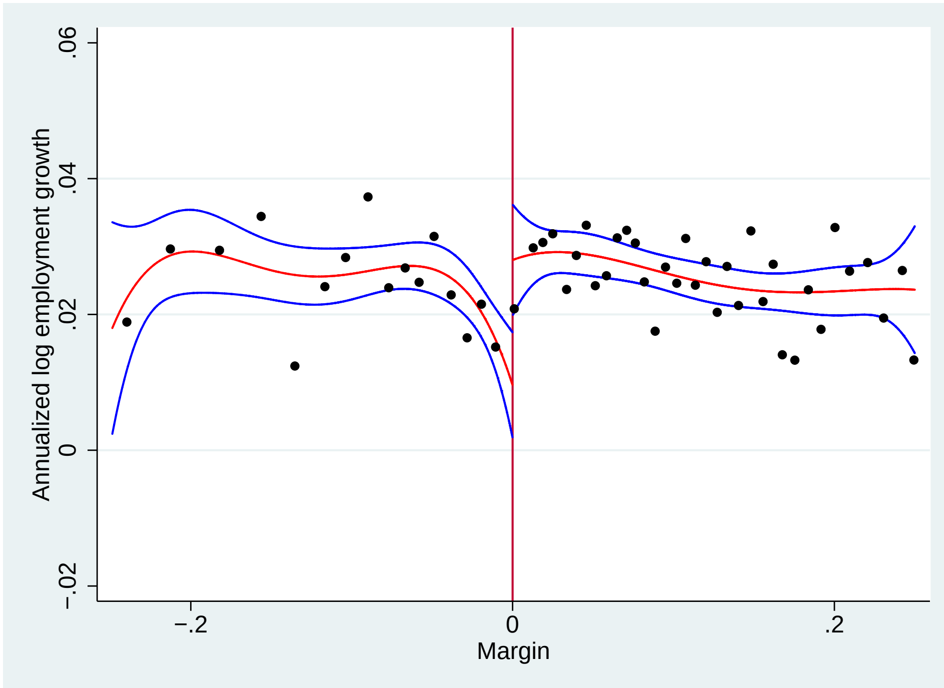


Figure 1.7: *Log employment growth vs. win margin of governing party candidate*

The figure plots the mean of log employment growth, in constituencies grouped by the win margin of the candidate representing the state-level governing party. Points to the right of zero indicate growth in locations won by the governing party candidate, while points to the left of zero indicate growth in locations lost by the governing party candidate. There are approximately fifty observations in each bin. A 4th degree polynomial function is fitted separately to each side of 0, and 95% confidence intervals are displayed.

value and 95% confidence interval of a 4th degree polynomial function fitted to the raw data, with separate specifications for aligned and non-aligned candidates. The jump in the regression line at zero is a visual analog of the estimates in Table 1.2.

We draw attention to three characteristics of this graph. First, the effect of alignment is large and significant when elections are close. Second, the effect of alignment appears to be highly local; constituencies won by a large margin do not grow employment at a different rate from those lost by a large margin. This finding is consistent with our conceptual framework: the returns to investing in economic outcomes have the highest electoral returns when elections are close.

Third, political alignment appears to have an effect primarily in constituencies lost by the aligned candidate. We emphasize that the regression discontinuity design does not identify this characteristic of the effect without further assumptions. There could be important unobserved variation between constituencies with close elections and those with wide victory margins.[28] Our empirical design allows us at best to make a causal claim only about the difference between aligned and non-aligned constiutencies. That said, the visual effect is striking, and we will discuss it in Section 1.6.

We now consider two outcomes over whose allocation we would expect politicians to have some influence: public sector employment and public goods.

Table 1.3 presents estimates of Equation 1.1, with employment disaggregated across private and public sectors. As above, estimates are reported (i) without controls; (ii) with lagged constituency controls, and (iii) with controls, and weighted with lagged baseline employment. The first three columns report estimates of the effect of political alignment on log public sector employment, which includes public sector establishments and state owned firms. The next three columns report the effect on log private sector employment. The coefficient of interest is almost twice as large for private firms as it is for government establishments, but the difference between the two is not statistically different from zero, based on a joint significance test. The effect of political alignment on employment in government firms is also indistinguishable from zero.[29] While we cannot say definitively that the effect of political alignment is driven by private sector firms, it is clear that politician influence over hiring at public sector firms is not what is driving the effect on total employment.

Tables 1.4 and 1.5 examine the effect of political alignment on local public infrastructure. Equation 1.1 is the basis for all columns of these tables, with different measures of constituency-level changes in public goods as the dependent variables.

Table 1.4 shows the RD estimates of politician alignment on construction of urban public

[28]One possibility is that close elections makes firms reluctant to invest because of uncertainty over the power of their political connections. If growth is lower when elections are close even in the absence of political intervention, then it is more difficult to determine whether the alignment effect is coming from aligned or non-aligned constituencies.

[29]The larger standard error is in part due to the fact that public employment is only 17-20% of total employment over the sample period.

Table 1.3: *Effect of politician alignment on log employment growth: Private sector vs. public sector*

	Public Sector Employment			Private Sector Employment		
Aligned (RD)	0.011	0.009	0.010	0.017	0.017	0.017
	(0.014)	(0.010)	(0.010)	(0.008)*	(0.007)**	(0.008)**
Margin	-0.320	-0.281	-0.325	-0.250	-0.225	-0.259
	(0.406)	(0.358)	(0.339)	(0.239)	(0.236)	(0.219)
Margin * Aligned	0.651	0.301	0.345	0.170	0.057	0.090
	(0.620)	(0.498)	(0.491)	(0.283)	(0.280)	(0.263)
Baseline Pub Emp		-0.066	-0.068			
		(0.015)***	(0.015)***			
Baseline Priv Emp					-0.024	-0.023
					(0.004)***	(0.004)***
Constant	-0.067	0.457	0.473	0.017	0.221	0.215
	(0.018)***	(0.119)***	(0.123)***	(0.010)	(0.034)***	(0.033)***
Weighted	No	No	Yes	No	No	Yes
Controls	No	Yes	Yes	No	Yes	Yes
N	662	662	662	663	663	663
r2	0.19	0.46	0.46	0.12	0.24	0.23

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. The table shows kernel regression discontinuity estimates of the effect of politician alignment with the state-level governing coalition on constituency log employment growth from 1990-98 and 1998-2005. The dependent variable in columns 1-3 is log employment growth in public sector firms. The dependent variable in columns 4-6 is log employment in private sector firms. In each group, the first column includes year and state fixed effects, the second adds lagged constituency controls controls, and the third weights observations by baseline employment. Standard errors are clustered at the state-election level.

infrastructure between 1991 and 2001. The dependent variables are aggregated from town data to the constituency level, and normalized by the mean and standard deviation in 1991. The columns respectively show the effect of alignment on: (1) km of paved urban roads; (2) number of urban electrical connections; (3) number of primary schools; (4) number of secondary schools; and (5) number of hospitals. None of the estimates are significantly different from zero. The standard errors indicate that, with 95% confidence, we can rule out positive effects in the range of 0.2-0.4 standard deviations from mean.

Table 1.5 shows the analogous table for rural public infrastructure. The columns are constituency aggregates of village data representing the: (1) share of villages with a paved access road; (2) share of villages with an electricity connection; (3) share of villages with a primary school; and (4) share of village land that is irrigated. As with the town data, none

Table 1.4: *Effect of politician alignment on urban public infrastructure*

	Roads	Electricity	Primary Schools	Secondary Schools	Hospitals
Aligned	-0.015	0.082	0.113	0.029	0.085
	(0.110)	(0.203)	(0.102)	(0.054)	(0.078)
Margin	-1.869	-0.800	3.852	-4.907	-3.800
	(3.214)	(7.708)	(5.624)	(3.455)	(1.207)***
Margin * Aligned	4.339	-1.125	-3.154	9.009	5.620
	(3.870)	(8.412)	(5.592)	(6.095)	(3.379)
Baseline	-0.027	-0.102	0.442	0.331	-0.084
	(0.064)	(0.178)	(0.189)**	(0.131)**	(0.304)
Constant	-0.000	0.297	-0.013	-0.009	-0.273
	(0.102)	(0.217)	(0.097)	(0.109)	(0.218)
N	465	465	465	465	465
r2	0.20	0.24	0.30	0.36	0.11

${}^{*}p < 0.10, {}^{**}p < 0.05, {}^{***}p < 0.01$. The table shows kernel regression discontinuity estimates of the effect of politician alignment with the state-level governing coalition on changes in the levels of local urban public infrastructure. The dependent variables have been normalized by the baseline level, so the coefficients can be interpreted as standard deviations. The dependent variables represent the following: (1) km of paved urban roads; (2) number of urban electrical connections; (3) number of primary schools; (4) number of secondary schools; and (5) number of hospitals. All regressions are run at the constituency level, with data aggregated up from individual towns. The data sources are the 1991 and 2001 Population Censuses. All regressions include state fixed effects. Standard errors are clustered at the state-election level.

of the estimates are significant. We can rule out effect sizes of 0.4 standard deviations for primary schools, and 0.2 standard deviations for the other measures.

In summary, we do not find a statistically significant effect of political alignment on either public sector hiring or construction of public infrastructure. The latter result is consistent with other work on India, which finds that citizen mobilization and national political agendas have played the dominant role in determining which regions gained public goods (Banerjee et al., 2005; Banerjee and Somanathan, 2007). The large negative coefficients on the baseline values in Table 1.5 suggest that this was a period where the least well-off villages experienced substantial growth in public infrastructure. We discuss in Section 1.6 discusses some other reasons that strategic allocation of public goods may not play a major role in closely contested constituencies.

Table 1.5: *Effect of politician alignment on rural public infrastructure*

	Roads	Electricity	Primary Schools	Irrigation
Aligned	0.036	0.033	0.161	-0.055
	(0.042)	(0.051)	(0.100)	(0.061)
Margin	-2.544	-1.792	1.404	-4.296
	(2.117)	(1.873)	(3.394)	(2.627)
Margin * Aligned	1.916	2.789	-5.672	4.111
	(2.123)	(2.428)	(4.690)	(3.881)
Baseline	-0.678	-0.481	-0.354	-0.255
	(0.109)***	(0.084)***	(0.055)***	(0.066)***
Constant	1.114	0.262	0.364	0.510
	(0.074)***	(0.099)**	(0.061)***	(0.129)***
N	460	460	461	443
r2	0.89	0.64	0.57	0.69

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. The table shows kernel regression discontinuity estimates of the effect of politician alignment with the state-level governing coalition on changes in the levels of local rural public infrastructure. The dependent variables have been normalized by the baseline level, so the coefficients can be interpreted as standard deviations. The dependent variables represent the: (1) share of villages with a paved access road; (2) share of villages with an electricity connection; (3) share of villages with a primary school; (4) share of villages with a hospital; and (5) share of village land that is irrigated. All regressions are run at the constituency level, with data aggregated up from individual villages. The data sources are the 1991 and 2001 Population Censuses. All regressions include state fixed effects. Standard errors are clustered at the state-election level.

1.5.2 Stock prices

Growth in private sector employment does not necessarily indicate that firms are better off. A firm could be worse off with increased employment if a politician is forcing the firm to hiring workers beyond the point of efficiency. To test whether market participants place a higher value on publicly traded firms when they are located in aligned constituencies, we examine the returns of firms after close elections between aligned and non-aligned candidates.

Table 1.6 report estimates from Equation 1.3, which identifies the effect of political alignment on the share prices of local firms. Column 1 is the baseline model without fixed effects. Column 2-4 respectively add fixed effects for (2) state; (3) state and year; and (4) state * year. The election of an aligned politician is associated with a positive abnormal return in

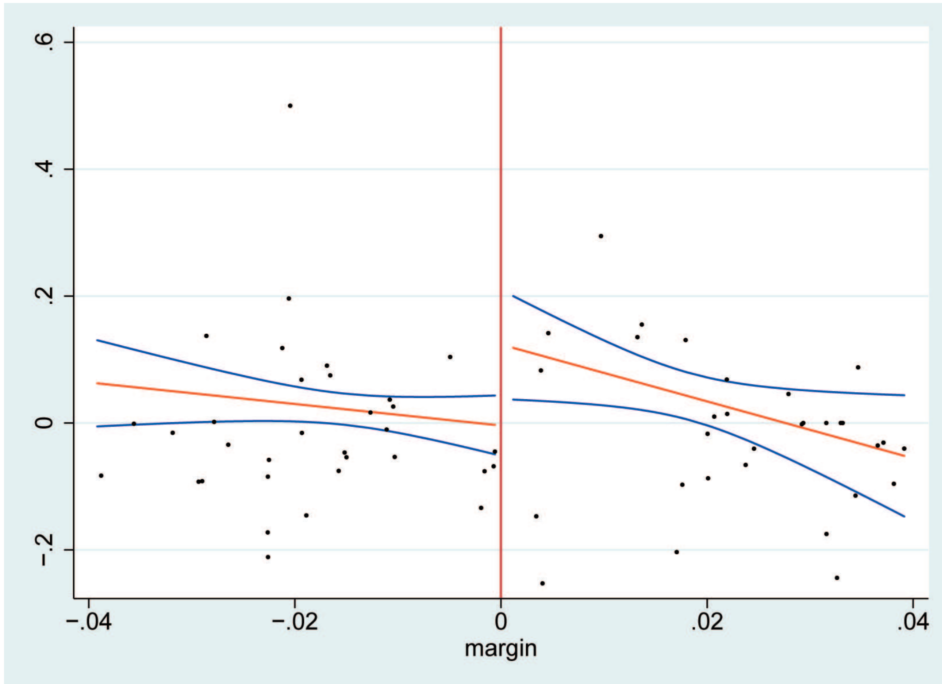


Figure 1.8: *Cumulative abnormal returns vs. win margin of governing party candidate*

The figure plots the mean of stock cumulative abnormal returns, adjusted for a market model, grouped by the win margin of the candidate representing the state-level governing party. Points to the right of zero indicate growth in locations won by the governing party candidate, while points to the left of zero indicate growth in locations lost by the governing party candidate. There are approximately twenty-five observations in each bin. A linear function is fitted separately to each side of 0, and 95% confidence intervals are displayed. The sample is all public traded firms with headquarters in towns not larger than a constituency.

the range of 12-15% in the month following the election. Figure 1.8 shows mean cumulative abnormal returns, sorted by the win or loss margin of the locally aligned candidate. Returns are visibly higher immediately on the positive side of the threshold.[30]

Columns 5 and 6 are placebo tests, using the cumulative abnormal return in the month before the election as the dependent variable. If election results are truly a surprise, we should identify no effect of election outcomes on pre-election returns. As expected, the coefficients are close to zero and not statistically significant.

[30] An alternate estimating approach would be to estimate the cumulative abnormal returns of firms, without focusing on the discontinuity. Even if a local election is not close, the uncertainty over which party becomes the majority party is resolved after an election. Appendix Table A.3 shows estimates from this test. Firms experience positive but smaller cumulative abnormal returns, consistent with the fact that less uncertainty is being resolved when elections are not close.

Table 1.6: *Effect of politician alignment on post-election stock returns*

	Event study				Placebo Test	
	(1)	(2)	(3)	(4)	(5)	(6)
Aligned	0.128	0.151	0.121	0.152	-0.014	0.021
	(0.060)**	(0.071)**	(0.077)	(0.090)*	(0.053)	(0.081)
Margin	-1.699	0.613	1.046	0.693	-0.108	-3.463
	(1.610)	(2.370)	(2.839)	(3.047)	(1.444)	(2.750)
Margin * Aligned	-2.787	-5.541	-5.431	-6.192	1.968	4.295
	(3.030)	(3.516)	(3.922)	(4.213)	(2.719)	(3.803)
Constant	-0.004	-0.139	-0.118	-0.086	0.009	-0.191
	(0.029)	(0.127)	(0.168)	(0.221)	(0.026)	(0.200)
Fixed Effects	None	State	State,Year	State * Year	None	State * Year
N	166	166	166	166	166	166
r2	0.03	0.21	0.35	0.36	0.01	0.34

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The table shows kernel regression discontinuity estimates of cumulative abnormal returns of publicly traded firms in the month following election. The independent variable *aligned* indicates that the winner of the constituency where the firm's headquarters are located is a member of the state-level governing coalition. Returns are measured against a market model with a value weighted index of Indian securities representing the market. Column 1 is the baseline model without fixed effects. Column 2 adds state fixed effects. Column 3 adds state and year fixed effects, and column 4 adds state-year fixed effects. Columns 5 and 6 conduct a placebo test, using the cumulative abnormal returns of firms in the month before the election as the dependent variable.

If politicians are forcing firms to hire excess labor for political reasons, as in Shleifer and Vishny (1994), we would expect political alignment to result in no effect or a negative effect on share price, as firms' labor decisions become sub-optimal. The positive effect rejects this explanation of the effect of political alignment on employment growth. In addition, it suggests market participants are informed about the importance of political alignment to firms, and price this information into stocks.

These results also corroborate our interpretation of the timing of results from Section 1.5.1. Our results on employment are based on total employment changes over a seven year period, during which other factors may have changed in unobservable ways. The stock price study is very precise on timing, showing that valuations of firms change in the precise month that an election takes place.[31]

1.5.3 Mechanisms

We next explore potential mechanisms that politicians could be using to affect local employment growth. We investigate three classes of mechanisms: (i) regulation; (ii) direct transfers; and (iii) supply of factors of production.

Table 1.7 estimates the standard specification, with an interaction variable specifying whether a firm is in an industry with a high dependence on bureaucratic officials. The respective measures used are industry-level measures generated from the following firm-level responses in international surveys: (1) Business was visited by government officials in last year; (2) Percentage of senior management's time spent dealing with government officials; (3) Needed an operating license in past year; and (4) Visited by tax officials in last year. The final column creates an index from the first principal component of an eigenvalue decomposition of the previous four variables. All interactions of *aligned*, *margin*, and the bureaucratic indexes are included in the regression, but not displayed for reasons of space. The coefficient of interest is the interaction between *aligned* and the high regulation dummy. The interaction coefficients in all five columns are positive and significant. Firms in industries that depend on bureaucratic inputs are more affected by the alignment of local politicians than firms in industries with less bureaucratic dependence. In fact, we detect no effect of the political variables on firms with low bureaucratic dependence: the uninteracted measure of political alignment is small and indistinguishable from zero. The larger sample size in this regression (and those following) reflects the fact that an observation is now a location-industry group pair, with the industry group defined by the interaction variable.

This result is consistent with the idea that politicians exert political influence through

[31] Note that publicly traded firms are considerably larger than the typical firm in the economic census (i.e. the sample for Table 1.2), and differ in many other ways. Finding results in both samples suggests that political factors affect a wide spectrum of firms. We cannot rule out that the effects are different for the two sets of firms. For example, it remains possible that politicians are forcing small firms to make inefficient hiring decisions, while giving other benefits to large firms that raise their valuation. We cannot test this, because we do not have valuations for small firms, nor high frequency employment for publicly traded firms.

Table 1.7: *Effect of politician alignment on log employment growth, interacted with dependence on bureaucratic inputs*

	(1)	(2)	(3)	(4)	(5)
Aligned (RD)	0.001	-0.006	-0.011	-0.002	-0.016
	(0.012)	(0.012)	(0.014)	(0.013)	(0.011)
Aligned * Visited by officials	0.023				
	(0.013)*				
Aligned * Mgmt time with officials		0.012			
		(0.007)*			
Aligned * Need operating license			0.016		
			(0.009)*		
Aligned * Visited by tax office				0.024	
				(0.011)**	
Aligned * Bureaucracy index					0.029
					(0.014)**
Un-interacted bureaucrat measure	0.018	0.056	0.053	0.046	0.036
	(0.011)	(0.015)***	(0.016)***	(0.013)***	(0.016)**
Constant	0.130	0.233	0.216	0.213	0.174
	(0.035)***	(0.029)***	(0.035)***	(0.032)***	(0.040)***
Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	State,Year	State,Year	State,Year	State,Year	State,Year
N	1326	1326	1326	1326	1326
r2	0.24	0.26	0.23	0.21	0.27

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. The table shows kernel regression discontinuity estimates of the effect of politician alignment with the state-level governing coalition on annualized log employment growth. The following industry-level measures of dependence of firms on government officials: (Column 1) Business was visited by government officials in last year; (2) % of Sr. mgmt time spent with officials; (3) Needed an operating license in past year; (4) Visited by tax officials in last year. Column 5 is an index consisting of the first principal component of an eigenvector decomposition of the previous four measures. The interaction between the bureaucrat measure and the *aligned* variable indicates the extent to which industries with dependence on bureaucrats are particularly affected by local political alignment. Margin, margin * aligned, and the interaction of these variables with the bureaucracy measures are included in the regression, but not displayed. Standard errors are clustered at the state-election level.

their control of the bureaucracy (Iyer and Mani, 2012), and have the ability to control the enforcement of regulation at a highly local level.[32] This result is discussed further in Section 1.6.

Procurement contracts to firms are a straightforward way for politicians to narrowly target government resources. The first two columns of Table 1.8 shows estimates from our primary specification, interacted with industry-level estimates of dependence on procurement. Column 1 uses the procurement measure from Indian data, while column 2 uses international data to classify industries. The interaction of alignment and procurement is not statistically different from zero in either test, and the effect of alignment remains positive and significant for low procurement industries. In combination with our null result on public sector employment, this suggests that direct transfers are not a major channel for the effect of political alignment.

Finally, in the class of factors of production, we consider the possibility that politicians are directing credit from state-owned banks to firms in party-aligned locations. Columns 3-6 in Table 1.8 shows RD estimates interacted with several industry measures of credit dependence and location measures of credit supply. None of the interaction terms of interest are statistically distinguishable from zero, and in most cases the point estimates are negative. Neither demand for credit nor availability of local banks significantly affects the relationship between political alignment and employment growth. An equally important class of factors of production normally supplied by the government is public infrastructure; our results above suggest that are not significantly affected by political alignment.

To control for other industry level factors and firm size, we test the interaction specification in a location-industry group level analysis. We use 2-digit ISIC groups, and apply the mean values of the bureaucratic dependence and credit indexes above, as well a dummy indicating that an industry tends to consist of small firms. Table 1.9 presents the estimates. The interaction of politician alignment and the regulatory index variable is positive and

[32] Some anecdotal examples of politician control over bureaucrats are described in Chaudury (2009) and The Hindu (2012).

Table 1.8: *Effect of politician alignment on log employment growth, interacted with measures of credit supply or demand, and procurement*

	Procurement		Credit			
	(1)	(2)	(3)	(4)	(5)	(6)
Aligned (RD)	0.019	0.016	-0.003	0.010	0.020	0.017
	(0.012)	(0.009)	(0.018)	(0.011)	(0.010)*	(0.009)*
Aligned * Procurement (India)	-0.012					
	(0.019)					
Aligned * Procurement (Int'l)		-0.003				
		(0.009)				
Aligned * New Loans			0.018			
			(0.015)			
Aligned * Finance demand				0.007		
				(0.013)		
Aligned * Public bank supply					-0.009	
					(0.014)	
Aligned * Credit Index						-0.007
						(0.011)
Constant	-0.130	0.016	0.012	-0.093	-0.014	-0.032
	(0.017)***	(0.015)	(0.028)	(0.015)***	(0.011)	(0.010)***
N	1321	1326	1325	1326	663	1326
r2	0.14	0.20	0.10	0.15	0.17	0.08

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. The table shows kernel regression discontinuity estimates of the effect of politician alignment with the state-level governing coalition on annualized log employment growth. Columns 1 and 2 interact all RD variables with measures of dependence on procurement. In each case, the coefficient of interest is the interaction between *aligned* and the industry measure. Column 1 uses Indian Enterprise Survey data, while column 2 uses international survey data. In columns 3-6, all RD variables are interacted with an industry-level measure of credit demand: (3) Demand for new loans; (4) Bank finance divided by working capital; (5) a location-level measure of the presence of local public sector banks; (6) a credit index consisting of the first principal component of an eigenvalue decomposition of all industry-level credit variables.

significant after controlling for firm size and industry dependence on credit.

1.6 Discussion

This section situates our findings in the context of the literature on state-business interaction in India, and considers possible models that can explain our results.

Our results are consistent with a model of politicians selecting between policy levers to maximize electoral advantage. We reconsider our three mechanisms in this light. Why would strategic politicians prefer regulation as a policy lever over direct transfers or construction of public infrastructure?

Public infrastructure is not well suited to strategic deployment in the closely contested constituencies studied in this paper. Public infrastructure is very costly, can take many years to build, and can be difficult to target to a single constituency. For example, an electricity line must cut across many constituencies, and would bring the clearest electoral advantage in a region where aligned constituencies are contiguous. If voter preferences are spatially correlated, clusters of aligned constituencies are less likely to have close elections.

Direct transfers to firms and individuals (we investigated procurement and public sector jobs) are more easily targeted, but they impose a financial cost to the government. Our results on procurement suggest either that alignment does not affect the allocation of government contracts, or that government contracts to firms do not have a measurable effect on overall employment. The sample period is a period of government disinvestment and privatization, so large expansions of public sector employment would be unlikely to take place.

Credit from state banks is also highly targetable. Lending in India is known to respond to political cycles (Cole, 2009), and there is evidence from other countries that politicians use state owned banks to reallocate private sector employment growth across legislative constituencies (Carvalho, 2010). Our lack of result on credit suggests either that this tool is too costly for politicians to use, or it does not have a substantial effect on overall employment, perhaps because politically directed lending tends to be unproductive (Khwaja

Table 1.9: *Location-industry group level RD Estimates of effect of majority alignment on employment growth*

	(1)	(2)	(3)	(4)	(5)
Aligned (RD)	0.004	-0.000	0.039	0.005	0.154
	(0.008)	(0.008)	(0.083)	(0.008)	(0.091)
Margin	-0.087	0.059	-1.569	-0.102	-5.000
	(0.283)	(0.308)	(3.809)	(0.305)	(3.355)
Margin * Aligned	0.378	0.210	0.990	0.300	2.769
	(0.388)	(0.468)	(7.449)	(0.468)	(6.559)
Bureaucrat index * Aligned	0.015			0.016	0.019
	(0.007)**			(0.006)**	(0.007)***
Credit index * Aligned		0.035		-0.009	-0.018
		(0.039)		(0.038)	(0.037)
Small * Aligned			-0.038		-0.163
			(0.091)		(0.101)
Bureaucrat index	0.001			0.050	-0.014
	(0.009)			(0.013)***	(0.008)*
Credit index		0.046		-0.334	-0.052
		(0.049)		(0.039)***	(0.035)
Small			0.427		0.908
			(0.258)		(0.157)***
Constant	-0.087	-0.095	-0.479	-0.045	-0.908
	(0.025)***	(0.023)***	(0.224)**	(0.026)*	(0.137)***
State, year fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
N	4326	4326	4326	4326	4326
r2	0.10	0.10	0.10	0.10	0.10

$^{*} p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$.

The table shows location-industry group level kernel regression discontinuity estimates of the effect of politician alignment with the state-level governing party on annualized log employment growth. The regression discontinuity variables *aligned*, *margin* and *margin* $*$ *aligned* are interacted with an index of industries dependence on bureaucrats, dependence on credit, and to have small mean establishment size. The interaction between the industry-level measures and the *aligned* variable indicates the extent to which that industry is particularly affected by local political alignment. Standard errors are clustered at the state-election-industry level.

and Mian, 2005).

Why do we find an effect on regulation? According to our conceptual framework, we should find that (i) local bureaucrats have the ability to significantly affect firm activity; and (ii) local politicians can control bureaucrats at low cost.

Bureaucrats in India affect firms through many channels. India is famous for its License Raj, the red tape intensive regime that exerted strict control over production and import decisions of firms through much of India's history. The 1990s saw significant reductions in licensing requirements, but by objective measures the Indian economy remained highly regulated throughout our sample period (Panagariya, 2008). The presence of red tape gives bureaucrats the ability to hold up formal sector firms that need operating licenses and permits.

Labor regulation is another domain where bureaucrats wield significant power over firms. India's 1947 Industrial Disputes Act requires companies above a certain size to seek government permission before firing any workers (Besley and Burgess, 2004). In practice, MLAs and bureaucrats play key roles as mediators in labor conflict in India. In addition to acting as bottlenecks when firms require government inputs, MLAs and bureaucrats have the ability to initiate tax audits and investigations; they can also control the intensity of investigations that have already begun.[33]

Is control over the bureaucratic process costly to politicians? Politicians have leverage over bureaucrats primarily through their ability to reassign them to less desirable posts (Iyer and Mani, 2012). This implicit threat means that it may cost nothing at all for a politician to control a bureaucrat. In equilibrium, the bureaucrat does a politician's will, and does not need to be reassigned. Enforcement of regulation also has a low political cost because it can be perceived as desirable by voters.

A final piece of evidence consistent with a regulatory channel is Figure 1.7, which shows that the effect of alignment on firms is concentrated in constituencies narrowly lost by the

[33] In the 2005 Enterprise Survey, 31% of firms disagreed with the statement "Government interpretations of regulation are predictable," and only 12% responded with "Fully agree."

ruling party. Our framework predicts that the ruling party puts equal value on increasing government inputs to aligned locations and decreasing inputs to non-aligned locations. Regulation is arguably a tool with an asymmetric effect: stricter enforcement of red tape can immediately halt economic activity, while more lenient enforcement may not have the same effect on increasing economic activity.[34] Even if politicians put equal effort into tightening regulation in non-aligned constituencies and loosening regulation in aligned constituencies, the economic effects may be greater in the areas where regulation is tightened.[35]

An alternate explanation for an effect of political alignment on employment growth is rooted in contracting. Suppose that local politicians play a key mediating role in the provision of government inputs, and that parties can only monitor politicians from their own party. The seminal red tape model of Banerjee (1997) predicts that the ruling party can improve the allocative efficiency of local services by increasing the level of red tape where agents are more difficult to monitor.[36] Under this model, red tape is increased in non-aligned constituencies, not due to electoral strategy, but as a second best allocative scheme.

This story is consistent with the finding that the effects of alignment are mediated through enforcement of regulation.[37] In the contracting model, however, it is not as clear why the effect of alignment on growth should be limited to swing constituencies, though as discussed above our empirical strategy does not provide clear inference on the importance of the swing effect.

---

[34]Consider a simple model in which a firm needs to pass multiple bureaucratic processes in order to operate. A barrier to any of these processes will halt firm operation. Removal of barrieres requires more coordination, as all constraints must be loosened before the firm can operate.

[35]Though politicians may also behave asymmetrically toward aligned and non-aligned places: Iyer and Mani (2012) finds that bureaucrats receive more post-election reassignments in non-aligned constituencies.

[36]Banerjee's model describes politicians and bureaucrats, but is equally applicable if local MLAs are central to the provision of government inputs.

[37]An alternate explanation with equivalent predictions is that bureaucrats are beholden to the party in power, and work less efficiently when monitored by unimportant non-aligned MLAs.

1.7 Conclusion

Firms in developing countries rely heavily on government inputs, access to which often depends on local politicians and bureaucrats. This paper draws on highly localized firm-level employment data to show that politician identity significantly affects firm growth in India.

Exploiting exogenous variation in politician identity induced by close elections, we show that the alignment of a local politician with the majority party at the state level strongly predicts increased private sector employment growth in the range of 1 to 2 percentage points per year. Further, in the month following elections, firms headquartered in aligned constituencies experience a 12-15% cumulative abnormal return, providing further evidence that political alignment is valuable to firms. However, politician alignment has no measurable effect on the supply of public infrastructure or public sector jobs.

The industries most affected by political alignment are those with a high dependence on bureaucratic inputs, and those likely to meet frequently with government officials. We hypothesize that within a constant state-level regulatory framework, the majority party can control where regulation is and is not enforced, giving the party significant control over local firms. This is consistent with evidence that politicians control bureaucrats through the threat of reassignment.

The evidence supports a model of rational politicians who take into account the costs and benefits of the different policy levers at their disposal. Regulatory discretion is a relatively low cost tool; in equilibrium, bureaucrats will be pliable even if no transfers take place. One major difference between the Indian context and that of studies that have found significant political effects on public infrastructure and public employment where we find none (Albouy, 2009; Cohen et al., 2011; Ferraz and Monteiro, 2010) could be that control over regulation gives Indian politicians a very cheap policy tool that is less easily manipulated in other countries.

Our empirical design does not identify whether the effect of political alignment is distortionary in the aggregate. While visual inspection of Figure 1.7 suggests that potential jobs are being destroyed in places narrowly lost by the majority party, we cannot determine whether these jobs are disappearing from the economy or shifting to other locations. However, the mechanism at play may be informative. Control over the intensity of regulatory enforcement is less likely to be distributional than control over public infrastructure. If politicians instruct bureaucrats in some locations to step up enforcement, this need not imply that enforcement in other locations is decreasing.[38] Public infrastructure is much more likely to be budget constrained, such that the construction of a road in one location may well preclude its construction in another.

India is well-known for its history of onerous regulation and barriers to doing business. While the high costs of adhering to regulation for Indian firms have been widely discussed, this paper sheds light on an additional cost of the regulatory state: if political control over regulatory enforcement is cheap, politicians may create additional distortions in pursuit of their electoral interests. This in turn provides an explanation for the persistence of high regulation in India: public officials may be reluctant to give up on a tool that affords them possibilities for rent extraction.

[38]This said, if the equilibrium value of regulation is too low, increases in regulation could be good for citizen welfare, even if they have a negative effect on employment. It is also possible that the party is increasing regulatory enforcement through assignment of stricter bureaucrats, in which case other locations might indeed experience less enforcement.

Chapter 2

Natural Resource Wealth, Mining Booms, and Local Economic Development in India[1]

2.1 Introduction

Does natural resource wealth put regions or nations on an adverse development path? Empirical work beginning with Sachs and Warner (1995) has led to the idea that the particular characteristics of wealth from natural resources could be a detriment to long-run growth, especially in countries with poor institutions.[2]

Most of the empirical evidence supporting a negative effect of natural resource wealth on economic development has been based on cross-country comparisons.[3] Recognizing the many confounding variables inherent to cross-country work, researchers have turned to

[1]Co-authored with Sam Asher.

[2]See Van der Ploeg (2011) for a survey of the literature.

[3]Sachs and Warner (1995) begins this literature. Mehlum et al. (2006) is a recent example, finding that natural resources a detrimental in the presence of bad institutions, but beneficial otherwise. Alexeev and Conrad (2009) is a dissenting voice in the cross-country literature, arguing that natural resources do not have adverse effects, but rather raise income without raising other indicators of development normally correlated with income.

within-country studies, and have focused on changes in resource extraction over time.[4]

This literature has found mixed results on the effect of natural resources on economic development. The interpretation of these mixed results is challenging, as there are many mechanisms by which natural resource wealth can affect local development, which include: (i) changes in local factor prices (Corden, 2012); (ii) local agglomeration spillovers (Ellison and Glaeser, 1997); (iii) increased public spending from mineral royalties; and (iv) deterioration in political outcomes (Robinson et al., 2006). This paper focuses on the first and second categories: the direct local economic consequences of mining booms. The context is India, a nation which as a whole is not highly dependent on point-source natural resources, but has many regions which produce coal, iron, gold and a range of other valuable minerals.

By using subnational and subregional variation, we are able to compare outcomes across regions that share basic political and economic institutions. We identify exogenous variation from changes in global prices of mineral resources found in India. We are also able to rule out the channel of government royalties, which are collected and spend at the state level in India.[5]

The cross sectional relationship between mineral deposits and economic structure are consistent with views that natural resource wealth can inhibit economic development, or set it onto a suboptimal growth path. Towns nearest mineral deposits are smaller, and controlling for their population, have higher employment in the mining sector, but significantly smaller manufacturing and retail trade sectors. They are also at higher altitude and further from political centers.

However, the geographic differences between mineral and non-mineral areas point to the risk of identifying the effects of natural resource wealth by comparing these types of locations. The locational choice of resource extraction industries is constrained by the

---

[4]Some recent examples include Carrington (1996), Aragon and Rud (2013), Caselli and Michaels (2013), and Domenech (2008).

[5]A lack of formal rule does not imply that royalties are not spent locally. However, neither public officials nor mining executives believed that public spending was being redirected toward communities affected by mining. This said, mining firms often make social investments in affected communities.

location of resources; towns with a high degree of natural resource extraction are likely to lack other natural advantages of resource-scarce locations, which could also produce a specialization pattern similar to that observed.

Exploiting time series variation in the value of mineral resources helps us understand the extent to which cross-sectional estimates are biased by unobserved characteristics of resource-rich regions. Contrary to the cross-sectional results, we find that exogenous increases in mineral resource wealth result in broad-based economic growth in nearby towns, across all manufacturing and service sectors. We find no evidence that growth in natural resource wealth results in a decline in sectors that compete for local factors of production.

Towns nearest mineral deposits experience the largest increase in employment growth, which is composed both of tradable and non-tradable sector growth. A 100% price shock to the value of a single nearby mineral deposit raises non-farm employment in the nearest towns by 17%, and manufacturing employment by 20%. The shock raises employment in the nearest villages by 9%, with larger increases in manufacturing and a decline in the regional rural service sector.

The difference between the long run characteristics of resource rich areas and the short term effects of mining booms suggests that some of the cross-sectional characteristics of mining areas are driven by characteristics of places other than their natural resources. In the short to medium run, our results contradict the class of theories that predict crowdout, and support a model of positive externalities from the natural resource extraction sector to nearby firms in other sectors.

Section 2.2 gives background information on the mining sector in India. Section 2.3 describes the key theoretical ways we should expect mining sector development to affect regional economies. Section 2.4 describes data sources, Section 2.5 describes the empirical strategy, and Section 2.6 presents results. Section 2.7 discusses interpretation of the results in the context of the theory, and concludes.

2.2 The mineral resource industry in India

Although modern India is not considered a mineral-rich country, it has a large and varied natural resource sector. In 2010, the mining sector employed 521,000 workers and accounted for 2.5% of national GDP (Indian Bureau of Mines, 2011). Mineral resources are unevenly distributed across India, and often make up a significant share of economic output in the places where they are concentrated. Over sixty different major minerals were mined in 2999 documented mines in 2010 (Indian Bureau of Mines, 2011).

Historically, Indian mines were predominantly state owned until significant privatization in the 1990s. By 2010, 2229 of 2999 mines were privately owned, representing 36% of total production value (Indian Bureau of Mines, 2011). Mineral deposits are found in nearly every state of India. The major exceptions are the Deccan Traps in west-central India and the highly populated states of Uttar Pradesh and Bihar in the lower Gangetic plain.

Major minerals such as iron ore are jointly regulated by the national and state governments, while minor minerals such as granite are regulated entirely by state governments. Notably, royalties and taxes paid by mining corporations are paid directly to state and federal governments. Importantly for this study, there is no requirement for fiscal proceeds from mining to be spent in communities close to mines. Further, our discussions with mining executives and public officials have not given any suggestion that these communities are likely to benefit disproportionately from spending of royalties or other public funds. However, mining companies often provide social goods, such as schools and libraries, to affected communities.

2.3 Conceptual Framework

This section considers possible channels by which mineral resource wealth could affect local industrial structure.

The first channel we consider is the Dutch Disease: in the presence of factor immobility, significant growth in one sector of the economy can increase the input factor prices faced by other sectors, making them less competitive (Corden, 2012). However, the new wealth associated with the booming sector increases the demand for locally produced goods. The factor cost effect will decrease the production of all sectors, while the demand effect increases the production of locally produced non-tradable goods. The Dutch Disease channel thus drives down the production of tradable goods, and could either increase or decrease the production of non-tradable goods.

A second channel is the agglomeration channel: growth in mineral extraction may have positive spillovers into other sectors (Ellison and Glaeser, 1997). The most widely discussed mechanisms originate in Marshall (1920): (i) input/output channels; (ii) thick labor markets; and (iii) knowledge spillovers. The first mechanism suggests that industries which have input/output linkages with the mining sector will benefit from mining booms. The second channel suggests that firms with similar labor forces to the mining sector will benefit.[6] We expect the last mechanism to be less important in the context of mining: mining sector knowledge is more likely produced in headquarter locations than at extraction sites. We also consider a fourth agglomeration channel: in a resource-poor country, many firms are constrained by a lack of basic infrastructure, such as roads and electricity, which market failures may prevent the private sector from providing. A local boom may increase the supply of government inputs, which other firms may then benefit from.[7]

Natural resource wealth could affect a local economy through a royalty channel, if local governments receive a share of profits from natural resource extraction, or if higher governments are required or choose to spend royalties in areas where mining takes place. We are able to largely exclude this channel, as there is no evidence that local communities or

[6] The short- and long-term mobility of labor are important here. If labor is immobile in the short-term, we would expect a mining boom to hurt sectors with similar labor demand to the mining sector, as wages are bid up. In the medium to long term, however, these firms would benefit from a thicker labor market, as labor moves into the area to work in the mining sector. It is worth noting that labor mobility in India tends to be very low (Foster and Rosenzweig, 2004).

[7] Two mechanisms for an increase in government inputs are possible. A mining boom may increase the value of local infrastructure, attracting efficiently allocated government inputs. Alternately, a booming industry may have increased lobbying power, and can thus be more effective at attracting government inputs to the area. Either way, we expect to see the largest growth in government inputs that are complementary to mining sector production.

governments receive a disproportion share of royalties from mining.[8] The absence of locally-spent royalties allows us to focus on the direct economic channels of mineral extraction, without the confounder of increased local government spending.

An additional widely discussed potential consequence of natural resource wealth is the behavior of political actors. As natural resource sectors are rent-rich, politicians may put more effort into appropriating some of that wealth. Corruption may increase, and potential entrepreneurs may move from organizing production to rent-seeking (Murphy et al., 1991)[9]. The economic effects of a political resource curse could include a fall in entrepreneurship (from the reallocation of entrepreneurial talent), and a deterioration in the quality of public goods (from the reallocation of politician effort). However, the lack of sector-specific predictions of a political channel make it difficult to test directly in our context.

Finally, natural advantage plays a major role in the economic characteristics of mineral rich regions. Economic centers arise in places with certain natural advantages, for example, in places well-suited for trade, such as ports or at the confluence of rivers. Mineral deposits are a kind of natural advantage that tends to be inversely correlated with other kinds of natural advantage: deposits are most often founded in highland areas that tend to be ill-suited for both trade and agriculture.[10] Unless mineral resources are positively correlated with other kinds of natural advantage, we should expect mineral-based economies to have fewer of the natural advantages of non-mineral economic centers. Cross-sectional correlations between natural resources and local economic structure are therefore importantly confounded by natural advantage. We eliminate confounding due to natural advantage by focusing on time series changes in the value of local mineral wealth. Natural advantages

[8]Mining companies do support local commnunity projects, such as schools and libraries. However, the scale of spending on these projects is much lower than royalties typically collected from mining projects.

[9]We explore some of these effects in parallel work, finding that elections are less competitive and criminal accusations against politicians are more likely when the local mining sector is booming (Asher and Novosad, 2013).

[10]Many valuable minerals are formed under pressure, deep in the Earth's crust. These minerals tend to be most accessible in mountainous areas, where geological activity has exposed these deeper layers.

other than those captured by the value of subsurface resources, are unlikely to significantly change over our sample period.

2.4 Data

The Indian Ministry of Statistics and Programme Implementation (MoSPI) conducted the 3rd, 4th and 5th Economic Censuses respectively in 1990, 1998 and 2005. The Economic Census is a complete enumeration of all economic establishments except those engaged in crop production and plantation; there is no minimum firm size, and both formal and informal establishments are included.

The Economic Census records information on the town or village of each establishment, whether ownership is public or private, the number and demographic characteristics of employees, the sources of electricity and finance, and the caste group of the owner. The main product of the firm is also coded using the 4-digit National Industrial Classification (NIC), which corresponds roughly to a 4-digit ISIC code. More detailed information on income or capital is not included. The main strengths of the data are its comprehensiveness, and rich detail on spatial location and industrial classification of firms.

We obtained location directories for the Economic Censuses, and then used a series of fuzzy matching algorithms to match villages and towns by name to the population censuses of 1991 and 2001.[11] We were able to match approximately 93% of villages between 1998 and 2005, and 81% from 1990 to 1998. The match rates for towns are respectively 78% and 55%. We also use data from the Population Census of India in 1991 and 2001, which includes village population and other demographic data, as well as information on local public infrastructure (roads, electricity, schools and hospitals).

Data on the location, type and size of mineral deposits come from the Mineral Atlas of India (Geological Survey of India, 2001), which provides the following characteristics of major mineral deposits in India: centroid latitude and longitude, mineral type, and

[11] The Economic Census of 1998 was conducted with the house listing for the 1991 population census, while the 2005 Economic Census used codes from the 2001 population census.

estimated reserves (in one of three size categories). Figure 2.1 shows a map of mineral deposit locations. Commodity prices come from the United States Geological Survey (Kelly and Matos, 2013). All prices are annual averages in the United States. Where available, we use the price for the ore as it is listed in the Indian deposit data. Where the ore price is unavailable, we match deposits to the price of the processed output of the mineral deposit (e.g. we use the price of aluminum for bauxite deposits).[12] We match deposits to villages and towns based on the geographic coordinates provided in the 2001 Population Census of India.

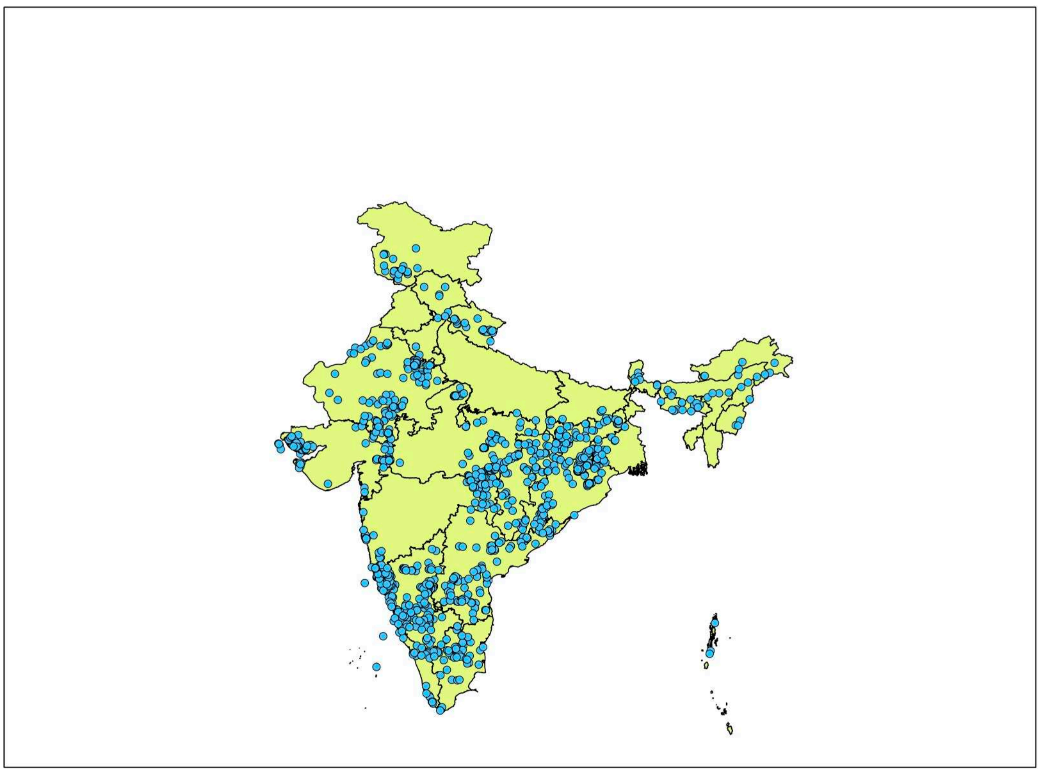


Figure 2.1: *Map of deposit locations*

Map of all major mineral deposits in India, as listed by the Mineral Atlas of India (Geological Survey of India, 2001).

The unit of observation is the village or town. For each location, we desire a measure

[12]Because we rely mainly on changes in prices over time, this imputation is reasonable as long as unprocessed and processed ore comove.

that indicates the extent to which the price of nearby mineral deposits has increased or decreased. There are two parts to this process: (i) creating a scalar measure that captures the recent price movement in a given commodity; and (ii) combining these price measures when locations are close to multiple deposits.

To capture recent changes in the value of a commodity, we use the mean price over the economic measurement period, and normalize it by the baseline price, measured at the beginning of the period. This measure is desirable in that a sustained increase in price results in a larger measured shock than a transitory increase in price. Further, a level shift in price at the beginning of the period results in a larger measured shock than a level shift in price at the end of the period, which is desirable if mineral price changes have lagged effects.

We use a 10-year trailing average for the baseline price, in order to prevent transitory shocks at the beginning of the measurement period from having too strong an effect on the price shock. The measure for a period of T years, ending in year t is given by Equation 2.1.

$$PriceShock_{c,t-T\rightarrow t} = \frac{\frac{1}{T}\sum_{\tau=t-T}^{t-1} p_{c,\tau}}{\frac{1}{10}\sum_{\tau=t-T-11}^{t-T-1} p_{c,\tau}}$$

Figure 2.2 presents mineral-wise price shocks for the period 1998-2005, the second of two periods in our sample.

For each location, we identify deposits within one of three concentric rings, with radii of 10km, 25km, and 50km. When multiple deposits fall within a zone, we take the sum of the price shocks.[13] Since commodity prices are increasing on average in the period, locations with more deposits will on average experience larger price shocks; we include a flexible function in the number of deposits near a location to control for this bias.

From a list of 45 minerals for which we have both deposit and price data, we discard economically unimportant minerals, defined as those for which the Indian Bureau of Mines does not publish production statistics or those whose average output per deposit is valued

[13]Our reasoning is that multiple deposits that are increasing in value create a larger shock to a local economy than a single deposit increasing in value. Results are largely robust to using a mean price shock. Future work will impute the total change in dollar value of reserves near a given location.

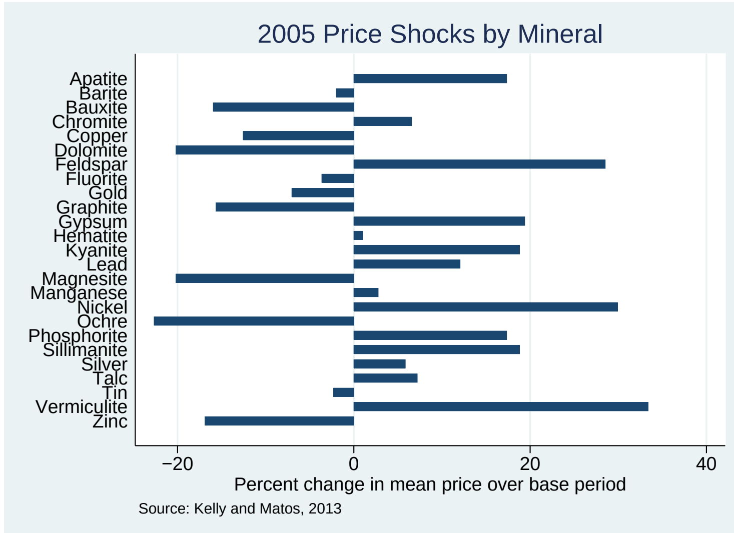


Figure 2.2: *Price shocks, 1998-2005*

at less than \$20,000 in 2005, the most recent year for which we have Economic Census data. We end up with 1325 deposits of 27 distinct minerals spread across 25 states in India. We use the presence of deposits rather than the presence of mines, as deposit existence is more likely to be exogenous than mine presence.[14] More common measures of resource abundance, such as share of GDP from primary commodities or mineral production value, are correlated with institutional factors and are better described as measures of resource dependence. Our use of mineral deposits avoids this endogeneity.

Table 2.1 shows key summary statistics separately for towns and villages.

[14]Known deposits are endogenous to the extent that their existence depends on some exploration having taken place. Future work will test robustness using a set of deposits that were known at the beginning of the sample period.

Table 2.1: *Summary Statistics*

Variable	Mean	StDev	N
Urban			
Non-farm employment (1990)	7872	19809	2678
Non-farm employment (1998)	10664	57544	4071
Non-farm employment (2005)	11142	41376	4063
Number of industries (2005)	82	33	4063
Mean firm size (2005)	2.75	2.93	4063
Paved road (km, 2001)	36.7	59.8	3705
Electrical connections (2001)	253	506	3739
Number deposits (10km)	0.3	1.1	3322
Number deposits (25km)	1.2	3.5	3322
Number deposits (50km)	3.8	6.8	3322
Number deposits (100km)	14.5	15.2	3322
Rural			
Non-farm employment (1990)	77	368	556939
Non-farm employment (1998)	87	287	588897
Non-farm employment (2005)	112	312	623289
Number of industries (2005)	10	10	622858
Mean firm size (2005)	2.01	7.69	623289
Paved approach (2001)	0.72	0.45	514606
Electricity (2001)	0.73	0.44	446655
Number deposits (10km)	0.2	0.9	510178
Number deposits (25km)	1.2	2.8	510178
Number deposits (50km)	4.4	6.9	510178
Number deposits (100km)	17.1	17.9	510178

2.5 Empirical strategy

The most common approach in the natural resource literature has been to regress the outcome variable on a measure of natural resource wealth, as in equation 3.1:[15]

$$Y_i = \beta_0 + \beta_1 * \text{RES}_i + \mathbf{1} * \mathbf{X_i'} + \epsilon_i, \tag{2.1}$$

where i indexes locations, RES_i is a measure of natural resource wealth, $\mathbf{X_i'}$ is a vector of location-specific controls and ϵ_i is an orthogonal error term.

This approach has had two major weaknesses. The first is with the measures of resource

[15]This is the approach used, among others, by Alexeev and Conrad (2009), Sachs and Warner (2001), Michaels (2010), Black et al. (2005) and Mehlum et al. (2006).

dependence used. Any measure with GDP in the denominator (for example, the often used primary export share of GDP) is subject to reverse causality: place that have failed to develop advanced sectors will necessarily have a high resource share of their economies. Measures of natural resource production are also endogenous: the extraction of natural resources may not take place if the background infrastructure and institutions are inadequate.

To escape the endogeneity of both mineral production and GDP, we proxy mineral resource wealth with the value of known mineral deposits, based on international prices. Geography is clearly exogenous to other factors.[16]

Estimating Equation 3.1 to compare resource-rich and resource-poor areas suffers from an additional omitted variable bias. Natural resources are not distributed at random; they are more likely to appear in regions that are mountainous and inaccessible, and as discussed in Section 2.3, resource-driven agglomerations are less likely to have other natural advantages due to compensating differentials. Control variables can mitigate this to some extent, but the high degree of selection on observables suggests that selection on unobservables may be significant as well (Altonji et al., 2005).

To better relate our work to the literature on cross-sectional estimation of the characteristics of resource rich places, we run the standard OLS tests. We use Equation 2.2:

$$Y_i = \beta_0 + \beta_1 * \mathrm{RES}_{10km,i} + \beta_2 * \mathrm{RES}_{25km,i} + \beta_3 * \mathrm{RES}_{50km,i} + \mathbf{1} * \mathbf{X_i'} + \epsilon_i, \tag{2.2}$$

where the multiple β coefficients capture the relationship between the economic outcome and the distance from the mineral deposit.

To eliminate omitted variable bias due to differences between mineral and non-mineral producing areas, we limit our study to mineral-rich areas, and rely on time series variation in the value of subsurface wealth, exogenously driven by international prices. We estimate Equation 2.3 to identify the effect of *changes* in mineral wealth on growth in the following

[16] Knowledge of mineral deposits remains endogenous, which may bias our OLS results. But this will not affect our time series results, which are limited to locations with known deposits, and locations with unknown deposits will not have mining sectors.

years:

$$Y_{i,t+1} - Y_{i,t} = \beta_0 + \beta_1 * pshock_{i,t} + \mathbf{1} * \mathbf{X_{i,t}} + \mathbf{fl_{s,t}} + \epsilon_{i,t}, \tag{2.3}$$

where $pshock_{i,t}$ is the change in value of nearby geological deposits as described above, $\mathbf{X_{i,t}}$ is a vector of location controls, $\mathbf{fl_{s,t}}$ is a state-year fixed effect and $\epsilon_{i,t}$ is an orthogonal error term. The coefficient β_1 identifies the effect of a change in mineral wealth on the economic outcome. As we are using interacted state-year fixed effects, our estimates are driven by variation in commodity price changes within a given state and time period. Standard errors are clustered at the district-level, which roughly corresponds to a shock radius of about 40km.

2.6 Results

2.6.1 Cross-section analysis

We begin by examining the cross-sectional relationship between subsurface mineral wealth and the economic characteristics of villages and towns. Equation 2.2 is the estimating equation. This method compares mineral rich to mineral poor regions within states, and estimates separate coefficients for places within 10km, 25km and 50km of mineral deposits. The reference group is composed of locations more than 50km from an economically valuable mineral deposit. The results should be interpreted as correlations between mineral wealth and economic outcomes, which are not necessarily causal.

Table 2.2 presents the cross-sectional relationship between mineral wealth, population and non-farm employment. Column 1 regresses village population on proximity to a mineral deposit, controlling for state fixed effects. In column 2, the dependent variable is non-farm employment, and population is an additional control. Villages are smaller in mineral rich regions, though this effect is muted in the villages closest to deposits. This is likely because deposits are located in remote regions of low population density, with agglomerations near the mines themselves.

Column 3 and 4 show the equivalent regressions for towns. Towns in mineral rich

Table 2.2: *Employment and demographic characteristics of mineral-rich regions*

	Pop (R)	Emp (R)	Pop (U)	Emp (U)
Deposit (10km)	-0.021	0.014	-0.062	-0.057
	(0.018)	(0.022)	(0.067)	(0.039)
Deposit (25km)	-0.045	-0.008	-0.069	0.003
	(0.018)**	(0.025)	(0.046)	(0.028)
Deposit (50km)	-0.116	0.028	-0.051	0.079
	(0.026)***	(0.027)	(0.048)	(0.028)***
Population (2001)		0.905		0.991
		(0.012)***		(0.014)***
Constant	5.914	-2.172	8.720	-1.036
	(0.078)***	(0.087)***	(0.127)***	(0.154)***
N	284744	284744	2798	2798
r2	0.16	0.52	0.15	0.83

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. The table show estimates from Equation 2.2. The model estimates a regression of log population or non-farm employment on a set of dummy variables indicating whether there is a major mineral deposit within 10km, 25km or 50km from the location. Columns 1 and 2 are at the village level, and columns 3 and 4 are at the town level. The dependent variable is log population (2001) in columns 1 and 3, and log non-farm employment (2005) in columns 2 and 4. All regressions include state fixed effects and standard errors are clustered at the district level.

regions are 5-6% smaller than towns in non-mineral regions; conditional on being within 50km of a mine, distance to the deposit has little additional effect. Controlling for population, employment is lower in towns nearest deposits, but 8% higher in towns that are 25-50km from a deposit. Mining operations are often headquartered in district capitals, which could explain these distance effects.

Table 2.3 presents estimates of Equation 2.2 on other characteristics of villages, controlling for population, non-farm employment and state fixed effects. Column 1 shows that villages nearest mines are likely to have higher employment in establishments with more than 50 employees.[17] Column 2 shows that villages nearest deposits have marginally more diverse non-farm economies, while villages in deposit regions but not directly on a deposit have less diverse economies. Columns 3 through 5 regress the existence of local public goods on

[17]We do not find a statistically significant relationship between deposits and average firm size. The vast majority of economic census firms are very small, so an unreasonably large increase in large firms would be required to substantially shift the mean size.

Table 2.3: *Other characteristics of mineral-rich villages*

	Emp (large)	Diversity	School	Paved road	Elect.	Dist. to town
Deposit (10km)	0.027	0.135	0.016	0.002	0.001	0.005
	(0.011)**	(0.080)*	(0.014)	(0.007)	(0.007)	(0.023)
Deposit (25km)	0.007	-0.039	0.047	0.000	-0.004	0.055
	(0.011)	(0.096)	(0.014)***	(0.006)	(0.006)	(0.020)***
Deposit (50km)	-0.024	-0.450	0.097	-0.009	-0.008	0.025
	(0.019)	(0.116)***	(0.023)***	(0.010)	(0.008)	(0.025)
Log employment (2005)	0.407	4.735	0.062	0.050	0.011	-0.055
	(0.018)***	(0.110)***	(0.007)***	(0.004)***	(0.002)***	(0.007)***
Population (2001)	-0.090	2.079	0.507	0.048	0.021	-0.033
	(0.011)***	(0.108)***	(0.016)***	(0.005)***	(0.004)***	(0.012)***
Constant	-0.448	-17.914	-2.289	0.139	0.847	3.162
	(0.052)***	(0.566)***	(0.110)***	(0.081)*	(0.024)***	(0.112)***
N	284744	284744	273281	232461	206993	278980
r2	0.19	0.69	0.32	0.33	0.70	0.13

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. The table shows estimates from Equation 2.2. The model estimates a regression of a village-level outcome variable on a set of dummy variables indicating whether there is a major mineral deposit within 10km, 25km or 50km from the village. The dependent variable associated with each column is: (i) Total non-farm employment in establishments with 25 or more employees; (ii) the number of distinct products produced by non-farm firms in the village; (iii) a dummy variable indicating the presence of a village primary school; (iv) a dummy variable indicating the village is accessible by a paved road; (v) a dummy variable indicating that the village is connected to the power grid; and (vi) the log of the distance to the nearest town in kilometers. All regressions include state fixed effects and standard errors are clustered at the district level.

the presence of mineral deposits. Villages in mining regions are 5-10% more likely to have primary schools, but equally likely to have a paved approach road or access to electricity. Column 6 shows that villages near mineral deposits are on average further from towns.[18]

Table 2.4 disaggregates rural employment effects by sector. The lack of relationship between mineral deposits and total employment masks significant differences in economic structure between mineral and non-mineral areas. Mining employment is higher. Villages nearest mines have smaller retail sectors, and higher employment in schools, health clinics and public administration. In addition to these larger sectors, villages in mining regions but further from actual deposits also have larger non-farm employment in agroprocessing and hotels and restaurants. These positive results are balanced in the total employment

[18]The relationship is not significant at the shortest distance, perhaps because a high density of mineral deposits may motivate a mining town.

regressions by the retail sector, which makes up a large share of rural non-farm employment. Non-agricultural manufacturing employment is largely unaffected by mining, but is a small part of many village economies. In summary, in the rural economy, mineral deposits are associated with increased employment in the mining and social sectors, and reduced employment in retail trade.

Table 2.5 shows the relationship between town characteristics and mineral deposits. Like villages, towns very close to mineral deposits have weakly higher employment in establishments with more than 50 employees. Industrial diversity is not affected by mineral deposits, nor are the number of primary schools, paved roads or electrical connections. However, towns in mining areas are considerably more remote; on average, they are 25-40% further from state capitals.

Table 2.6 describes the relationship between mineral deposits and the sectoral composition of non-farm employment in towns. As with villages, the similarities in overall employment numbers mask sectoral differences. Towns nearest to mineral deposits have 30% higher employment in the mining sector, but significantly lower employment in manufacturing and retail (respectively 16% and 12%). These effects decay as distance grows, and towns located 25-50km from mines have higher employment, controlling for population, in a broad spectrum of industries - these are likely district capitals.

Towns nearest mineral deposits exhibit some classic characteristics of Dutch Disease: enlarged resource sectors and diminished manufacturing sectors. The next section exploits time series variation in the value of mineral resources to shed light on whether this result is causal or driven by unobserved characteristics of resource-rich regions.

2.6.2 Time series analysis

Table 2.7 shows estimates from Equation 2.3, which identify the effect of exogenous changes in mineral resource wealth on non-farm employment growth in nearby towns and villages. The independent variable (Price Shock) is the average price of local minerals over the period between census measurements, normalized by the 10-year moving average of the price at

Table 2.4: *Economic structure of mineral-rich villages*

	Mining	Ag Proc	Constr.	Manuf.	Retail	Ed&Health	Hotels	Community	Gov.
Deposit (10km)	0.031	0.011	0.004	0.017	-0.053	0.030	0.053	0.006	0.030
	(0.007)***	(0.054)	(0.008)	(0.026)	(0.022)**	(0.017)*	(0.021)**	(0.026)	(0.017)*
Deposit (25km)	0.015	0.037	-0.003	-0.012	-0.053	0.036	0.026	-0.013	0.009
	(0.005)***	(0.059)	(0.009)	(0.027)	(0.020)**	(0.019)*	(0.020)	(0.028)	(0.019)
Deposit (50km)	0.034	0.187	0.017	0.014	-0.001	0.100	0.134	-0.018	0.080
	(0.005)***	(0.070)***	(0.010)	(0.042)	(0.033)	(0.029)***	(0.027)***	(0.036)	(0.031)**
Population (2001)	0.051	0.421	0.106	0.862	0.901	0.646	0.505	0.576	0.434
	(0.003)***	(0.030)***	(0.007)***	(0.022)***	(0.013)***	(0.017)***	(0.017)***	(0.017)***	(0.015)***
Constant	-0.306	-1.877	-0.586	-3.954	-3.949	-2.630	-2.840	-2.884	-2.340
	(0.020)***	(0.194)***	(0.042)***	(0.166)***	(0.096)***	(0.112)***	(0.112)***	(0.108)***	(0.092)***
N	284744	284744	284744	284744	284744	284744	284744	284744	284744
r2	0.01	0.06	0.03	0.30	0.46	0.29	0.20	0.24	0.17

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. The table shows estimates from Equation 2.2. The model estimates a regression of village-level employment in a given non-farm sector on a set of dummy variables indicating whether there is a major mineral deposit within 10km, 25km or 50km from the village. The sectors are (1) mining; (2) agroprocessing; (3) construction; (4) non-agricultural manufacturing; (5) retail trade; (6) education and health; (7) hotels and restaurants; (8) community organizations; and (9) public administration. All regressions include state fixed effects and standard errors are clustered at the district level.

Table 2.5: *Other characteristics of mineral-rich towns*

	Emp (large)	Diversity	Schools	Km paved	Elect. Conn.	Dist. to Cap.
Deposit (10km)	0.097	-0.030	0.047	-0.027	0.021	0.076
	(0.150)	(1.080)	(0.040)	(0.050)	(0.041)	(0.053)
Deposit (25km)	0.156	0.659	0.038	-0.038	0.037	0.258
	(0.110)	(0.855)	(0.032)	(0.031)	(0.028)	(0.072)***
Deposit (50km)	-0.004	-0.697	0.001	0.012	-0.035	0.370
	(0.092)	(0.926)	(0.029)	(0.035)	(0.024)	(0.104)***
Log employment (2005)	2.243	17.663	0.113	0.143	0.280	0.047
	(0.091)***	(0.882)***	(0.029)***	(0.035)***	(0.032)***	(0.049)
Population (2001)	-0.467	8.575	0.710	0.803	0.686	-0.153
	(0.100)***	(1.057)***	(0.033)***	(0.039)***	(0.037)***	(0.059)**
Constant	-8.724	-149.208	-5.302	-6.144	-0.296	5.543
	(0.458)***	(4.200)***	(0.172)***	(0.145)***	(0.129)**	(0.359)***
N	2798	2798	2726	2569	2563	2797
r2	0.62	0.82	0.71	0.72	0.83	0.34

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. The table show estimates from Equation 2.2. The model estimates a regression of a town-level outcome variable on a set of dummy variables indicating whether there is a major mineral deposit within 10km, 25km or 50km from the town. The dependent variable associated with each column is: (i) Total non-farm employment in establishments with 50 or more employees; (ii) the number of distinct products produced in the town; (iii) the number of primary schools in the town; (iv) the kilometers of paved road in the town; (v) the number of electricity connections in the town; and (vi) the log of the distance to the state headquarter town. All regressions include state fixed effects and standard errors are clustered at the district level.

the beginning of the period. Interacted time / state fixed effects are included, and standard errors are clustered at the district level, which corresponds to the approximate range we believe price shocks have a direct effect on the local economy.

Columns 1 and 2 show estimates on employment growth in villages. Increases in the value of local mineral resources have positive and significant effects on non-farm employment in the nearest villages. Limiting the sample to villages within 10km of a mineral deposit (column 1), a doubling in the value of a single mineral deposit results in a 9% increase in total employment. To examine the spatial dimension of this effect, we widen the sample to locations within 50km of a mineral deposit, and allow separate coefficients for price changes in minerals at different distances from the deposit. Column 2 indicates that the effect is highly local: the effect is concentrated in villages nearest the mineral deposit; villages 10-50km from deposits show only a statistically insignificant estimate of 2%.

Table 2.6: *Economic structure of mineral-rich towns*

	Mining	Constr.	Manuf.	Retail	Ed&Health	Hotels	Community	Gov.
Deposit (10km)	0.322	0.094	-0.165	-0.118	-0.102	0.017	-0.051	-0.005
	(0.130)**	(0.123)	(0.079)**	(0.051)**	(0.078)	(0.067)	(0.060)	(0.134)
Deposit (25km)	0.304	0.142	-0.084	-0.074	-0.012	0.129	-0.019	-0.017
	(0.121)**	(0.094)	(0.062)	(0.035)**	(0.053)	(0.051)**	(0.042)	(0.094)
Deposit (50km)	0.231	0.317	0.085	0.035	0.080	0.232	0.076	0.219
	(0.095)**	(0.093)***	(0.056)	(0.035)	(0.048)*	(0.053)***	(0.043)*	(0.107)**
Population (2001)	0.651	1.076	1.014	1.009	0.996	0.962	0.987	1.126
	(0.029)***	(0.032)***	(0.022)***	(0.017)***	(0.021)***	(0.023)***	(0.017)***	(0.039)***
Constant	-5.465	-8.624	-3.620	-3.273	-4.292	-4.785	-4.932	-6.163
	(0.298)***	(0.330)***	(0.242)***	(0.186)***	(0.231)***	(0.248)***	(0.184)***	(0.436)***
N	2798	2798	2798	2798	2798	2798	2798	2798
r2	0.19	0.36	0.63	0.79	0.67	0.59	0.72	0.39

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. The table shows estimates from Equation 2.2. The model estimates a regression of town-level employment in a given non-farm sector on a set of dummy variables indicating whether there is a major mineral deposit within 10km, 25km or 50km from the village. The sectors are (1) mining; (2) agroprocessing; (3) construction; (4) non-agricultural manufacturing; (5) retail trade; (6) education and health; (7) hotels and restaurants; (8) community organizations; and (9) public administration. All regressions include state fixed effects and standard errors are clustered at the district level.

Table 2.7: *Time series effects of mineral wealth on rural/urban non-farm employment growth*

	Rural	Rural	Urban	Urban
Price shock (10km)	0.088	0.063	0.167	0.128
	(0.047)*	(0.032)*	(0.090)*	(0.097)
Price shock (25km)		0.022		0.057
		(0.022)		(0.054)
Price shock (50km)		0.020		0.046
		(0.017)		(0.026)*
Baseline	-0.441	-0.433	-0.423	-0.459
	(0.010)***	(0.007)***	(0.066)***	(0.038)***
Log population	0.421	0.409	0.440	0.464
	(0.013)***	(0.010)***	(0.076)***	(0.043)***
Constant	-0.955	-0.916	-0.383	-0.356
	(0.074)***	(0.053)***	(0.234)	(0.116)***
State*Year F.E.	Yes	Yes	No	No
N	41255	255725	451	2289
r2	0.24	0.23	0.37	0.32

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. The table shows estimates from Equation 2.3. The dependent variable is rural employment growth in columns 1 and 2, and urban growth in columns 3 and 4. The time period is the seven or eight years between economic censuses; growth figures are not annualized. The sample in columns 1 and 3 is limited to locations within 10km of a mineral deposit. Columns 2 and 4 extend this distance to 50km and have separate coefficients for price shocks at these greater distances, to capture broader spatial effects. All regressions include state-time fixed effects and standard errors are clustered at the district-period level.

Columns 3 and 4 show estimates in towns. Limiting the sample to the on average 225 towns per period within 10km of a mineral deposit (Column 3), we find that a doubling in the value of a single nearby deposit results in a 17% increase in non-farm employment. When we estimate the distance gradient (column 4), the effect is more dispersed: the point estimate remains largest for towns nearest the deposit, but towns from 25-50km of a mineral deposit also show a 5% increase in employment.

We next separate these employment effects into manufacturing and service sector employment. Table 2.8 shows estimates for villages. Columns 1 and 3 show that the overall employment increase in nearby villages can be decomposed into a loss in service sector employment and an increase in manufacturing, the reverse prediction of a Dutch Disease model. Columns 2 and 4 investigate the distance gradient. The increase in manufacturing jobs is located nearest the mineral deposits, while the loss in service sector jobs affects a wider region, extending up to 50km from the mineral deposit.[19] We speculate that the

[19]The zero coefficient on the 0-10km price shock variable does not necessarily imply no loss in service sector

Table 2.8: *Rural Sector Groups*

	Manufacturing	Manufacturing	Service	Service
Price shock (10km)	0.037	0.073	-0.058	0.008
	(0.069)	(0.058)	(0.027)**	(0.021)
Price shock (25km)		0.010		-0.018
		(0.032)		(0.014)
Price shock (50km)		-0.025		-0.052
		(0.025)		(0.012)***
baseline	-0.453	-0.433	-0.504	-0.512
	(0.018)***	(0.013)***	(0.011)***	(0.009)***
Log population	0.282	0.269	0.458	0.460
	(0.015)***	(0.012)***	(0.016)***	(0.011)***
Constant	-0.731	-0.762	-1.124	-1.133
	(0.086)***	(0.070)***	(0.077)***	(0.051)***
N	33025	205694	40159	248642
r2	0.25	0.24	0.29	0.28

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. The table shows estimates from Equation 2.3. The dependent variable is sector-level rural employment growth. Columns 1 and 2 look at manufacturing sector employment, and columns 3 and 4 look at service sector employment. The time period is the seven or eight years between economic censuses; growth figures are not annualized. The sample in columns 1 and 3 is limited to locations within 10km of a mineral deposit. Columns 2 and 4 extend this distance to 50km and have separate coefficients for price shocks at these greater distances, to capture broader spatial effects. All regressions include state-time fixed effects and standard errors are clustered at the district-period level.

villages nearest mineral deposits are supplying inputs to the mine. The loss in service sector jobs can be explained if rural retail employment is a low productivity diversification strategy. If employment in a mine is preferable to retail sector employment, we would see people move away from the retail sector when mine hiring increases.

Table 2.9 decomposes sectors further, roughly following 2-digit ISIC codes. The decomposition of manufacturing into food- and non-food manufacturing shows that manufacturing growth is mainly in agro-processing, with little effect in other manufacturing sectors. This is not surprising, if villages have a comparative advantage in processed agricultural goods relative to other manufactured goods. The decline in services is relatively broad, affecting retail trade, education and health and public administration. Retail trade is the most important of these as a share of rural employment; however, the decline in government and social services may have important welfare effects. An exception to the decline in

jobs in the villages nearest to the deposit. Villages nearest the deposits are also likely to be in the 25-50km range of other similar deposits. A zero coefficient therefore indicates no additional effect of being within 10km of a deposit, conditional on being 25-50km from a deposit.

village service employment is an increase in employment in community services, which include libraries, museums and religious and community organizations. These are likely services funded directly by the mining sector for purposes of corporate social responsibility; baseline employment in community services is very small, so the growth in this sector is not economically substantive.

Table 2.10 decomposes resource-driven urban growth into manufacturing and service sectors. Effects are largest in the manufacturing sector, which grows 19-21% in response to a doubling in price of a single nearby deposit. Point estimates suggest service sector growth in the range of 10-16%, though it is not statistically significant. While estimates for service and manufacturing growth are not statistically different from each other, we can rule out the notion that non-tradable growth is crowding out the tradable sector. As above, the distance gradient shows an effect weakly declining in distance.

Table 2.11 further decomposes the sectoral effects of natural resource wealth. As with the rural sector, the positive effect on manufacturing is concentrated in the processing of agricultural commodities. The effect on non-agricultural tradable products is positive but insignificant. Service sector growth is broad-based, with growth in employment in construction, retail trade, education, health and public administration. In short, resource booms appear to result in broad based growth across all urban sectors. Non-agricultural manufacturing growth is somewhat lagging, with the smallest point estimate of all sectors, but is nevertheless consistently positive and close to 10% across all specifications.

2.7 Conclusion

We provide new evidence on the relationship between local natural resource wealth and economic development. We improve upon previous studies of local resource effects by concentrating on highly local effects within a large, poor country. Our economic estimates are most significant in places within 10km of mineral deposits, suggesting that much of the work on natural resources may be overlooking the highly localized impact of natural resource extraction. We compare causal estimates of resource wealth to the cross-sectional

Table 2.9: *Time series effects of mineral wealth on economic structure of nearest villages*

	Ag Proc	Non-ag Manuf.	Constr.	Retail	Ed&Health	Hotels	Community	Gov.
Price shock (10km)	0.142	-0.057	0.031	-0.110	-0.098	-0.008	0.113	-0.087
	(0.089)	(0.049)	(0.019)	(0.034)***	(0.043)**	(0.037)	(0.045)**	(0.049)*
Baseline sector employment	-0.419	-0.454	-0.692	-0.531	-0.553	-0.427	-0.617	-0.524
	(0.029)***	(0.013)***	(0.018)***	(0.014)***	(0.013)***	(0.013)***	(0.016)***	(0.012)***
Log population	0.123	0.395	0.100	0.499	0.388	0.264	0.357	0.274
	(0.014)***	(0.018)***	(0.008)***	(0.018)***	(0.015)***	(0.011)***	(0.015)***	(0.013)***
Constant	-0.507	-1.612	-0.496	-1.964	-1.117	-1.188	-1.735	-1.161
	(0.087)***	(0.117)***	(0.064)***	(0.109)***	(0.077)***	(0.074)***	(0.089)***	(0.074)***
N	41255	41255	41255	41255	41255	41255	41255	41255
r2	0.25	0.27	0.41	0.30	0.31	0.20	0.34	0.31

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01.$

	Ag Proc	Non-ag Manuf.	Constr.	Retail	Ed&Health	Hotels	Community	Gov.
Price shock (10km)	0.101	0.008	0.003	0.010	-0.015	0.032	0.048	-0.020
	(0.070)	(0.039)	(0.015)	(0.025)	(0.027)	(0.029)	(0.029)	(0.032)
Price shock (25km)	0.063	-0.029	0.018	-0.038	-0.029	0.010	0.045	-0.029
	(0.043)	(0.025)	(0.011)*	(0.017)**	(0.021)	(0.018)	(0.021)**	(0.020)
Price shock (50km)	-0.006	-0.037	0.019	-0.097	-0.060	-0.041	0.042	-0.040
	(0.028)	(0.021)*	(0.007)***	(0.012)***	(0.019)***	(0.016)***	(0.018)**	(0.016)**
Baseline sector employment	-0.390	-0.444	-0.720	-0.532	-0.559	-0.438	-0.623	-0.537
	(0.020)***	(0.010)***	(0.010)***	(0.009)***	(0.010)***	(0.009)***	(0.011)***	(0.011)***
Log population	0.126	0.388	0.094	0.497	0.382	0.252	0.364	0.252
	(0.013)***	(0.014)***	(0.005)***	(0.012)***	(0.010)***	(0.008)***	(0.011)***	(0.010)***
Constant	-0.595	-1.614	-0.481	-1.969	-1.047	-1.124	-1.789	-1.025
	(0.076)***	(0.088)***	(0.034)***	(0.073)***	(0.052)***	(0.054)***	(0.063)***	(0.058)***
N	255725	255725	255725	255725	255725	255725	255725	255725
r2	0.24	0.26	0.42	0.30	0.31	0.21	0.35	0.31

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01.$ The table shows estimates from Equation 2.3. The dependent variable is sector-level rural employment growth. The top panel limits the sample to locations within 10km of a mineral deposit, while the bottom panel includes locations up to 50km from a mineral deposit, with separate coefficients from shocks to minerals at different distances from locations. The sector used as dependent variables is: (1) agroprocessing; (2) non-ag manufacturing; (3) construction; (4) retail trade; (5) education and health; (6) hotels and restaurants; (7) community organizations; and (8) public administration. The time period is the seven or eight years between economic censuses; growth figures are not annualized. All regressions include state-time fixed effects and standard errors are clustered at the district-period level.

Table 2.10: *Urban Sector Groups*

	Manufacturing	Manufacturing	Service	Service
Price shock (10km)	0.213	0.190	0.161	0.104
	(0.083)**	(0.090)**	(0.101)	(0.109)
Price shock (25km)		0.101		0.075
		(0.076)		(0.063)
Price shock (50km)		-0.016		0.036
		(0.039)		(0.034)
baseline	-0.381	-0.383	-0.464	-0.464
	(0.022)***	(0.022)***	(0.022)***	(0.022)***
Log population	0.373	0.374	0.488	0.488
	(0.025)***	(0.025)***	(0.024)***	(0.024)***
Constant	-1.064	-1.085	-0.695	-0.694
	(0.119)***	(0.118)***	(0.091)***	(0.092)***
N	3963	3963	3962	3962
r2	0.26	0.26	0.30	0.30

$^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$. The table shows estimates from Equation 2.3. The dependent variable is sector-level urban employment growth. Columns 1 and 2 look at manufacturing sector employment, and columns 3 and 4 look at service sector employment. The time period is the seven or eight years between economic censuses; growth figures are not annualized. The sample in columns 1 and 3 is limited to locations within 10km of a mineral deposit. Columns 2 and 4 extend this distance to 50km and have separate coefficients for price shocks at these greater distances, to capture broader spatial effects. All regressions include state-time fixed effects and standard errors are clustered at the district-period level.

relationship between mineral wealth and development that has been most commonly studied, allowing us to demonstrate the bias in the latter studies.

The cross sectional relationship between mineral deposits and economic structure are consistent with views that natural resource wealth can inhibit economic development, or set it onto a suboptimal growth path. Towns nearest mineral deposits are smaller, and controlling for their population, have higher employment in the mining sector, but significantly smaller manufacturing and retail trade sectors.

Exploiting exogenous variation in the value of mineral resources helps us understand the extent to which cross-sectional estimates are biased by unobserved characteristics of resource-rich regions. Contrary to the cross-sectional results, we find that exogenous increases in mineral resource wealth result in broad-based economic growth in nearby towns, across all manufacturing and service sectors. We find no evidence that growth in natural resource wealth results in a decline in sectors that compete for local factors of production.

Table 2.11: *Time series effects of mineral wealth on economic structure of nearest towns*

	Ag Proc	Non-ag Manuf.	Constr.	Retail	Ed&Health	Hotels	Community	Gov.
Price shock (10km)	0.435	0.102	0.356	0.205	0.117	0.092	0.213	0.252
	(0.217)**	(0.101)	(0.185)*	(0.103)**	(0.125)	(0.137)	(0.150)	(0.118)**
Baseline sector employment	-0.511	-0.330	-0.563	-0.384	-0.470	-0.396	-0.572	-0.380
	(0.033)***	(0.042)***	(0.032)***	(0.050)***	(0.061)***	(0.049)***	(0.046)***	(0.041)***
Log population	0.519	0.357	0.642	0.427	0.563	0.416	0.592	0.540
	(0.054)***	(0.063)***	(0.051)***	(0.054)***	(0.068)***	(0.056)***	(0.056)***	(0.052)***
Constant	-3.661	-1.339	-4.254	-1.215	-2.050	-1.391	-2.625	-2.439
	(0.484)***	(0.364)***	(0.633)***	(0.248)***	(0.327)***	(0.299)***	(0.344)***	(0.316)***
N	451	451	451	451	451	451	451	451
r2	0.36	0.23	0.44	0.28	0.34	0.26	0.34	0.35

$^{*}p < 0.10,^{**}\ p < 0.05,^{***}\ p < 0.01.$

	Ag Proc	Non-ag Manuf.	Constr.	Retail	Ed&Health	Hotels	Community	Gov.
Price shock (10km)	0.365	0.085	0.191	0.170	0.149	0.077	0.132	0.208
	(0.219)*	(0.101)	(0.170)	(0.102)*	(0.121)	(0.128)	(0.142)	(0.144)
Price shock (25km)	-0.014	0.107	0.025	0.009	0.050	0.000	0.049	-0.026
	(0.122)	(0.073)	(0.124)	(0.055)	(0.075)	(0.071)	(0.072)	(0.130)
Price shock (50km)	-0.017	-0.027	0.167	0.058	-0.055	0.053	0.086	0.087
	(0.075)	(0.040)	(0.070)**	(0.034)*	(0.043)	(0.049)	(0.046)*	(0.053)
Baseline sector employment	-0.512	-0.348	-0.574	-0.370	-0.507	-0.388	-0.570	-0.383
	(0.017)***	(0.028)***	(0.015)***	(0.030)***	(0.025)***	(0.023)***	(0.023)***	(0.016)***
Log population	0.443	0.339	0.708	0.397	0.539	0.409	0.570	0.497
	(0.030)***	(0.033)***	(0.028)***	(0.031)***	(0.030)***	(0.026)***	(0.026)***	(0.025)***
Constant	-3.004	-1.024	-5.093	-0.958	-1.597	-1.492	-2.382	-1.975
	(0.291)***	(0.170)***	(0.264)***	(0.129)***	(0.140)***	(0.143)***	(0.155)***	(0.176)***
N	2289	2289	2289	2289	2289	2289	2289	2289
r2	0.37	0.23	0.44	0.24	0.38	0.24	0.35	0.31

$^{*}p < 0.10,^{**}\ p < 0.05,^{***}\ p < 0.01.$ The table shows estimates from Equation 2.3. The dependent variable is sector-level urban employment growth. The top panel limits the sample to locations within 10km of a mineral deposit, while the bottom panel includes locations up to 50km from a mineral deposit, with separate coefficients from shocks to minerals at different distances from locations. The sector used as dependent variables is: (1) agroprocessing; (2) non-ag manufacturing; (3) construction; (4) retail trade; (5) education and health; (6) hotels and restaurants; (7) community organizations; and (8) public administration. The time period is the seven or eight years between economic censuses; growth figures are not annualized. All regressions include state-time fixed effects and standard errors are clustered at the district-period level.

Our results are measured at a 7- or 8-year time horizon, and thus identify medium term effects of resource wealth on local employment growth. It remains possible, but in our view, unlikely, that natural resource wealth has an impact on the manufacturing sector that is positive in the short term, but negative in the long term. Future work will attempt to identify the longer-term effects of natural resource wealth in India.

The fixed geographic characteristics of local natural resources make it challenging to identify a causal effect of resource wealth. Our results point to the importance of focusing on evidence that is plausibly causal, as our cross-sectional results point to a very different story from our causal time series results. In future work, we plan to investigate further dimensions of the relationship between resource wealth and economic development that have been widely discussed. In particular, several studies have found that natural resource are beneficial in democratic and well-governed economies like Norway, but detrimental in autocratic and poorly-governed economies like those of many sub-Saharan countries.

India is country with a consolidated democracy, but a wide range of governance quality across its many states and territories. The very high quality of India's democracy, controlling for its level of wealth, may explain the positive effects of natural resource wealth that we have found. Exploiting variation in the quality of governance across India allows us to test whether the quality of governance has an important impact on the role of natural resource in economic development.

Chapter 3

The Employment Effects of Rural Road Construction in India[1]

3.1 Introduction

Universal access to paved roads, much like clean water and consistent electricity, remains an unreached goal in many developing countries, particularly in rural areas. Fifty-four years after independence, 33% of Indian villages did not have a paved approach road in 2001 (Population Census). The absence of such infrastructure raises trade costs and reduces access to both outside markets and government services. The high costs of infrastructure investments mean that both economic and political considerations tend to guide their placement, posing challenges for the estimation of their impact. In this paper we exploit the allocation rules of a large-scale rural road construction program in India to estimate the impact of feeder roads on rural nonfarm economic activity.

The Pradhan Mantri Gram Sadak Yojana (PMGSY) – the Prime Minister's Village Road Program – was launched in 2000 with the goal of providing all-weather access to unconnected habitations across India. The government developed specific guidelines to prioritize large, unconnected habitations: those with populations above 1000 were to receive highest priority,

[1]Co-authored with Sam Asher.

followed by those with populations above 500. Lower priority was given to smaller localities and those with existing "fair-weather" roads. At the start, about 170,000 habitations were eligible for the program, a number that has grown as the guidelines have been expanded to include smaller habitations. By March 2011, over 420,000 km of roads had been sanctioned to connect nearly 110,000 habitations at a cost of 1.19 trillion INR ($27 billion) (Ministry of Rural Development, 2012).

These rules provide us with three distinct ways to estimate the impact of a new road on rural economic activity. First, we provide OLS estimates of the relationship between PMGSY road construction and employment growth, based on the timing of road construction in villages that eventually received roads. However, endogenous timing of road construction creates unknown bias in these estimates. We address this in two ways. Program rules create discontinuities in the likelihood of receiving a road at populations of 500 and 1000, allowing us to use a fuzzy regression discontinuity approach to estimate the impact. Finally, we take advantage of the fact that planning and implementation was carried out at the district level. As a result, the probability of receiving a road early in the program was a function not only of village size, but also of its relative size within its district. We are thus able to instrument for road treatment with the within-district population rank, controlling flexibly for population.

We construct a new dataset that combines data on road construction with village characteristics and economic outcomes. We match Population Census data (1991, 2001) to Economic Census data (1998, 2005) to measure the economic consequences of road construction during the first five years of the PMGSY (2000-04). The Economic Census is a complete enumeration of nonfarm economic establishments in India, covering over 4000 towns and 500,000 villages. It contains, among other variables, data on employment and industrial sector for each establishment, allowing us to estimate the effect of rural roads on employment growth, firm size and formality.

We interpret the construction of a village feeder road under the PMGSY as a large reduction in the costs of moving goods, capital and people to and from a village. Theoretically, it is unclear what impact such a change in factor mobility should have on village economic activity. For example, out-migration may increase as the costs of travel decrease, or fall as economic opportunity expands in the village. Nonfarm employment may grow as firms serving outside demand now face sufficiently low transportation costs to be competitive in outside markets; alternatively, villages may further specialize in agriculture as trade for outside goods becomes more feasible. Increasing returns to scale, such as those due to fixed costs, would predict that certain firms may only exist when transport costs are low enough to enable them to access markets beyond the confines of a village. Thus road construction may not result in the specialization predicted by simple trade models but rather diversification of the village economy as it becomes integrated into a larger trade network.

We find that the construction of a road results in a significant increase in nonfarm village employment. While our results vary depending on the specification, all estimates predict over 100 percent growth in nonfarm employment upon receiving a road. As we assign treatment based on road completion, we interpret these not as temporary increases due to ongoing construction but medium-term changes in the level of village employment. In future work, we will test whether these results represent a change in the level or growth rate of village employment, a distinction recently explored in research on the economic impact of railroads in China (Banerjee et al., 2012a).

In addition to aggregate nonfarm employment, we find evidence that roads facilitate significant structural transformation and diversification of the local economy. Average village firm size increases, as does the number of industries present in the village. These findings are consistent with evidence from Nepal suggesting that villages closer to cities serve much larger markets, with greater specialization of household economic activities but greater diversification at the village level (Fafchamps and Shilpi, 2005). Further, we provide evidence that growth is significantly higher in villages that also have a supply of electricity, suggesting significant complementarities between transportation and power infrastructure.

We also provide some of the first well-identified estimates of the cost effectiveness of rural road construction. We find that, assuming no positive or negative employment spillovers to non-treated villages, \$1,000,000 in rural road construction generates between 500 and 700 nonfarm jobs, or approximately \$1400 to \$2000 per job. Given that India's per capita GDP in 2005 was \$732, our findings suggest very high returns to rural road construction, a conclusion supported by other attempts to estimate the returns to infrastructure investment (Fan and Hazell, 2001). These results should be interpreted with caution, as we do not observe concurrent changes to agricultural employment, and thus are unable to estimate changes in total village employment.

The rest of the paper proceeds as follows: Section 3.2 summarizes the most relevant literature on transport costs and rural roads. Section 3.3 provides a description of the PMGSY rural road construction program. Section 3.4 describes our empirical strategies. Section 3.5 explains the data used. Section 3.6 presents results and discussion. Section 3.7 concludes.

3.2 Literature

Recent years have seen a renewed interest in the importance of transportation costs in facilitating growth and development, particularly in the trade literature. Limao and Venables (2001) use quotes from a shipping company and other sources of data to estimate the impact of inter-country transportation costs on trade flows, concluding that much of the low trade volume in sub-Saharan Africa is due to high transportation costs that result from poor quality infrastructure. Djankov et al. (2010) estimate that every additional day required to ship a container between two countries is associated with a reduction in bilateral trade of more than 1%, in addition to causing a distortion in trade away from time-sensitive exports. Intra-national transport costs exacerbate the challenge of realizing gains from trade; Atkin and Donaldson (2012) estimate that internal trade costs in Ethiopia and Nigeria are 7-15 times larger than in the United States, greatly reducing the benefits of globalization.

Infrastructure has long been one of the priorities of economic development policy and research: fully 15% of World Bank spending between 1995 and 2005 was dedicated to infrastructure projects, with 42% of that amount spent in China and India alone (The World

Bank, 2007). Recent research has utilized novel identification strategies to investigate the link between the expansion of infrastructure and local economic performance. Banerjee et al. (2012a) examine the impact of railroads in China, finding that while railroads caused a level increase in income, nearby locations grew no faster than farther locations during a nearly 20 year period of rapid economic growth. Storeygard (2012) interacts global oil price shocks with distances to the nearest port to investigate the impact of transport costs on urban economic activity in Africa, finding a significant inverse relationship between transport costs and urban economic output, as proxied by nighttime luminosity. Donaldson (2012) develops a multi-region, multi-commodity trade model to assess the impact of railroad construction in colonial India, estimating that the expansion of the railroad network into a region increased real income by approximately 16% and greatly reduced trade costs. Michaels (2008) finds that the construction of the US Interstate Highway System generates sectoral and wage growth consistent with standard trade theory.

Another strand of research estimates the impact of changes in market access rather than average transportation costs generally. Redding and Sturm (2008) find that cities close to the border between West and East Germany experienced population loss relative to cities further from the border whose market access was less impeded by the partition of Germany following World War II. Hornbeck and Donaldson (2012) seek to unite the market access literature with the infrastructure literature by estimating the impact of railroads on agricultural land values using a market access approach, finding that the railroad network had by 1890 more than tripled the total value of agricultural land in the United States.

Of course, there are many reasons why rural roads may have very different economic effects when compared to core infrastructure projects such as interstate highways and long-distance railroads. Casaburi et al. (2012) use a fuzzy regression discontinuity design to examine the effect of village feeder roads on agricultural markets, finding that such roads significantly lower market prices of local agricultural goods. Gollin and Rogerson (2010) build a multi-sector model and calibrate it using data from Uganda to understand the relationship between agricultural productivity, transport costs and economic activity. They predict that high transport costs produce inefficiently high specialization in agriculture, and that investments in road infrastructure would lead to significant reallocation of labor to the nonfarm economy. Khandker et al. (2009) use propensity score matching to evaluate the impact of a rural road program in Bangladesh, estimating that receiving a road lowers village poverty by 5-6% while increasing household consumption by 8-10%, although subsequent work suggests that some of these gains may not persist over time (Khandker and Koolwal, 2011). Most closely related to this paper, Banerjee et al. (2012b) study the impact of the PMGSY on a broad range of outcomes in a sample of 267 villages in Uttar Pradesh. They find that road construction results in greater access to government services, lower consumer prices, higher agricultural prices, increased employment outside of agriculture and less daily migration, with no effect on longer-term migration. While the majority of evidence points towards large economic gains from the construction of rural roads, some studies have suggested that low incomes and population densities in rural areas may not generate sufficient demand for transportation services on rural roads (Raballand et al., 2011).

Methodologically, our use of within-district population rank is similar to Andrabi et al. (2013), who use local population rank to instrument for the placement of a girls' secondary school. Banerjee et al. (2012b) also use within-district population rank to estimate the effects of the PMGSY although, as discussed above, their focus is on the response of households, rather than firms, to the construction of a new road.

3.3 Context and background

The Pradhan Mantri Gram Sadak Yojana (PMGSY) – the Prime Minister's Village Road Program – was launched in 2000 with the goal of providing all-weather access to unconnected habitations across India. The focus was on the provision of new feeder roads to localities that did not have access, although in practice many projects under the scheme upgraded pre-existing roads. Originally, the stated goal was to provide all habitations with populations greater than 1000 with connectivity by 2003 and all habitations with population greater than 500 with connectivity by 2007. These thresholds were to be lower in desert and tribal areas, as well as hilly states and districts affected by left-wing extremism.[2]

Although funded and overseen by the federal Ministry of Rural Development, responsibility for road construction was delegated to state governments. District Rural Road Plans were drafted for every district in India. Funding comes by a combination of taxes on diesel fuel (0.75 INR per liter), central government support and loans from the Asian Development Bank and World Bank. By March 2011, over 420,000 km of roads had been sanctioned to connect nearly 110,000 habitations at a cost of 1.19 trillion INR ($27 billion) (Ministry of Rural Development, 2012).[3] The mandate of the program has recently been expanded to include all habitations with populations above 100.

3.4 Empirical Strategy

Our goal is to estimate the effect of the construction of a new rural road on changes in village-level economic activity. We start by estimating the OLS relationship between road construction and village employment growth using the following estimating equation:

$$Y_{v,s,t} = \beta_0 + \beta_1 * newroad_{v,s} + \zeta X_{v,s,t-1} + \eta_s + \epsilon_{v,t-1}, \tag{3.1}$$

where $Y_{v,s,t}$ is log employment in village v in state s at time t, and *newroad* indicates that a new feeder road was constructed in village v at time t, $X_{v,s,t-1}$ is a vector of village controls measured at baseline, and η_s is a state fixed effect.

The problem with this approach is that roads are not allocated at random. Roads are expensive infrastructural investments likely to be demanded by nearly all unconnected villages. Governments may target such investments to locations that are particularly needy, or have high potential for growth, or are politically connected or favored by powerful politicians. Controlling for baseline village characteristics or regional fixed effects will

[2]Habitations are defined as clusters of population whose location does not change over time. They are distinct from, but form parts of, revenue villages used by the Economic and Population Censuses. See National Rural Roads Development Agency (2005) for more details.

[3]We use an exchange rate of 44.06 INR per USD, the average for 2005

Table 3.1: *Tabulation of Villages Receiving PMGSY Roads by Year*

	Year Completed										
Year Sanctioned	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**Total**
2000	6	166	609	1080	990	604	315	419	111	61	4361
2001	0	0	21	720	1387	1372	704	316	161	126	4807
2003	0	0	0	2	403	1251	1153	682	364	266	4121
2004	0	0	0	0	0	300	774	791	861	474	3200
2005	0	0	0	0	0	0	178	1108	1541	1839	4666
2006	0	0	0	0	0	0	1	569	1047	2263	3880
2007	0	0	0	0	0	0	0	0	134	1150	1284
2008	0	0	0	0	0	0	0	0	0	149	149
Total	6	166	630	1802	2780	3527	3125	3885	4219	6328	26468

reduce this bias, but it remains a concern at the heart of the literature on infrastructure.

We undertake three strategies to mitigate bias from selection in the placement of roads: (i) using OLS to exploit variation in the timing of PMGSY road construction; (ii) using regression discontinuity to exploit population thresholds that determine road priority; (iii) using an instrumental variable approach to exploit within-district population ranking of villages, which is an additional determinant of road prioritization.

3.4.1 Road timing (OLS)

The naive OLS is biased because the types of villages that receive new rural roads differ on many unobserved characteristics from the types of villages that do not receive new roads. By limiting our sample to villages that eventually did receive PMGSY roads, and exploiting the timing of road construction, we can eliminate any confounders that differ between villages that did and did not receive new roads.

Table 3.1 shows the number of rural roads built under the PMGSY in each year. Of the 75,399 roads in our sample that were built by 2009, 25,354 were built by 2005, the year in which we measure employment outcomes.[4] We then estimate Equation 3.1 on this limited sample, where the treatment variable indicates that a road was built before 2005.

Under the assumption that the order in which roads were constructed under the PMGSY

[4]The 2005 Economic Census was conducted from late 2005 to early 2006.

is uncorrelated with other factors that affect growth, this would be an unbiased estimate of the effect of rural roads on village outcomes. However, this assumption is tenuous: order of construction is likely to be influenced by both political and economic factors that could bias OLS estimates either upwards or downwards.

3.4.2 Population priority thresholds (RD)

Timing of road construction is endogenous, which means the OLS estimates of the effects of roads are likely to be biased. In order to overcome this endogeneity, we exploit the population thresholds intended to guide the allocation of roads under the PMGSY. State implementing officials were instructed to target habitations in the following order: (i) habitations with population greater than 1000; (ii) habitations with populations greater than 500; and lastly, (iii) habitations with populations greater than 250.

Even if selection into PMGSY treatment is biased by political or economic factors, these factors are not likely to change discontinuously at these population thresholds. If these rules were followed to any degree by state officials, the likelihood of PMGSY treatment will discontinuously increase at these population thresholds, making it possible to estimate the effect of the program using a fuzzy regression discontinuity design.

Under the assumption of continuity at the treatment threshold, the fuzzy RD estimator (Imbens and Lemieux, 2008) estimates the local average treatment effect (LATE) of receiving a new road, for a village with population equal to the threshold:

$$\tau = \frac{\lim_{pop \to T^+} \mathbb{E}[Y_v | pop_v = T] - \lim_{pop \to T^-} \mathbb{E}[Y_v | pop_v = T]}{\lim_{pop \to T^+} \mathbb{E}[newroad_v | pop_v = T] - \lim_{pop \to T^-} \mathbb{E}[newroad_v | pop_v = T]}, \tag{3.2}$$

where pop_i is habitation population, T is the threshold population, and $newroad_i$ is an indicator variable for whether village v received a new road in the sample period.

A given population threshold increases the probability of receiving a road by a different amount in different states. For example, states with a large number of large, unconnected villages, are more likely to have large first stages at the high threshold of 1,000. Analysis of PMGSY documentation and discussions with public officials have led us to focus on the population threshold of 500, in five major states which cover a large share of the geographic and economic diversity of India: Gujarat, Chhattisgarh, Assam, Maharashtra and Tamil Nadu.

We estimate the reduced form fuzzy RD using equation 3.3:

$$Y_{v,s,t} = \beta_0 + \beta_1 (pop_{v,s,t-1} > T) + \beta_2 pop_{v,s,t-1} + \beta_3 pop_{v,s,t-1} * (pop_{v,s,t-1} > T) + \zeta X_{v,s,t-1} + \eta_s + \epsilon_{v,s,t}, \tag{3.3}$$

where $Y_{v,s,t}$ is log village employment at time t, T is the population threshold of 500, $pop_{v,s,t}$ is habitation population at time t, $X_{v,s,t-1}$ is a vector of village controls measured at baseline, and η_s is a state fixed effect. Village controls and state fixed effects are not necessary for identification but improve the efficiency of the estimation. The local average treatment effect of a road, identified in a village at the population threshold T, is $\beta_1 + \beta_3 * T$. For ease of exposition, we subtract the threshold value 500 from the population variable, such that $T = 0$, and β_1 fully describes the treatment effect.

The fuzzy regression discontinuity approach accurately identifies the treatment effect of rural roads, under the assumption that crossing the population threshold affects the probability of receiving a road, and nothing else of significance. There are two potential threats to this identification strategy. First, if other village characteristics vary discontinuously at the threshold in a way that we are unable to control for (e.g. if participation in other government programs uses the same thresholds), then our estimates will be biased. Second, if the running variable (habitation population) can be manipulated, randomness of assignment at the threshold is violated. We discuss this possibility when we describe results from the regression discontinuity strategy below.

3.4.3 Population rank (IV)

In addition to the population threshold rules, district-level planning and implementation of the PMGSY meant that priorization was determined not only by population but also by relative population ranking within a district: a village would receive higher prioritization than an equivalent village if it had fewer larger eligible villages in its district. Holding population constant, a village in a district with many larger unconnected villages is less likely to receive a new road under the PMGSY. Under the assumption that, after controlling flexibly for total population, the population rank of a village within a district does not affect a village's growth prospects except through the likelihood of receiving a road through PMGSY, instrumental variable estimation provides an unbiased estimate of the effect of a rural road (Angrist and Lavy, 1999).

Our empirical specification is:

$$Y_{v,s} = \beta_0 + \beta_1 * newroad_{v,s} + \beta_2 f(pop_{v,s}) + \zeta X_{v,s} + \eta_s + \epsilon_{v,s} \tag{3.4}$$

where $Y_{v,s}$ is the outcome of interest in village v in state s, $newroad_{v,s}$ is an indicator for whether the village received a road under the PMGSY, $f(pop_{v,s})$ is a function of village population, $X_{v,s}$ is a vector of village controls and η_s is a state fixed effect. We estimate Equation 3.4 using $RANK_{v,s}$, the within-district population rank of village v, as an instrument for $newroad_{v,s}$. In alternate specifications, in order to reduce noise, we instrument for $newroad_{v,s}$ with a dummy variable indicating that $RANK_{v,s} < 75$.[5]

The estimation provides an unbiased estimate of the effect of a new rural road on employment growth, so long as the exclusion restriction is not violated: $RANK_{v,d,t-1}$ must affect growth only through the increased likelihood of obtaining a new road under the PMGSY. In Section 3.6 we discuss robustness checks to ensure satisfaction of the exclusion restriction.

3.5 Data

3.5.1 PMGSY

Data on the PMGSY is generated through the Online Management and Monitoring System (OMMS), the software used in program tracking and implementation. These data are not a survey - they are the administrative records of the actual program. Data include but are not

[5] Our results are robust to different cutoffs for this dummy variable and are available upon request.

limited to road sanctioning and completion dates, cost and time overruns, contractor names, and quality monitoring reports.

PMGSY data are reported at either the habitation or the road level. There is a many-to-many correspondence between habitations and roads: roads serve multiple habitations, and habitations may be connected to multiple roads. Habitations are subsets of census villages, which tend to comprise between one and three habitations; approximately 200,000 villages consist of only a single habitation.

3.5.2 Economic and population census

The Indian Ministry of Statistics and Programme Implementation (MoSPI) conducted the 4th and 5th Economic Censuses respectively in 1998 and 2005.[6] The Economic Census is a complete enumeration of all economic establishments except those engaged in crop production and plantation; there is no minimum firm size, and both formal and informal establishments are included.

The Economic Census records information on the town or village of each establishment, whether ownership is public or private, the number and demographic characteristics of employees, the sources of electricity and finance, and the caste group of the owner. The main product of the firm is also coded using the 4-digit National Industrial Classification (NIC), which corresponds roughly to a 4-digit ISIC code. More detailed information on income or capital is not included. The main strengths of the data are its comprehensiveness, and rich detail on spatial location and industrial classification of firms.

We obtained location directories for the Economic Censuses, and then used a series of fuzzy matching algorithms to match villages and towns by name to the population censuses of 1991 and 2001.[7] We were able to match approximately 93% of villages between 1998 and 2005. We also use data from the Population Census of India in 1991 and 2001, which

[6] The 6th Economic Census is ongoing at the beginning of 2013.

[7] The Economic Census of 1998 was conducted with the house listing for the 1991 population census, while the 2005 Economic Census used codes from the 2001 population census.

includes village population and other demographic data, as well as information on local public infrastructure (roads, electricity, schools and hospitals).

We matched PMGSY data to economic and population census data at the village level, using population census codes where they were reported in the PMGSY, and a Hindi-language fuzzy matching algorithm to match village names across the two datasets. We successfully matched over 85% of habitations listed in the PMGSY to their corresponding population census villages.

Table 3.2 shows village-level summary statistics for the entire sample of villages used in our analysis.

Table 3.2: *Summary statistics*

Variable	Mean	(Std. Dev.)	N
New road	0.049	(0.216)	181232
Employment (1998)	68.079	(100.469)	181232
Employment (2005)	84.257	(119.934)	181232
Ln employment growth	0.213	(0.876)	181232
Firm count (1998)	33.719	(45.697)	181232
Firm count (2005)	45.368	(60.037)	181232
Ln firm count growth	0.289	(0.842)	181232
2001 Population	1422.329	(1021.174)	181232
Pop growth 1991-2001	1.202	(0.263)	181232
Irrigation share	0.431	(0.365)	174649
Ln land area	5.363	(1.085)	174649
Distance from town	20.832	(19.453)	180783
Diversity (1998)	8.727	(6.503)	181232

3.6 Results

3.6.1 OLS

Table 3.3 presents OLS estimates of the relationship between log employment growth (1998-2005) and treatment, defined as having received a completed PMGSY road by 2005. The sample is all locations that received a PMGSY road before 2012. Column 1 presents the estimate only controlling for 1998 (log) employment and village population. Column 2 introduces state fixed effects. Column 3 introduces standard village level controls of share of land irrigated, log land area, distance from

Table 3.3: *OLS: Employment growth on roads*

	(1)	(2)	(3)	(4)
New road before 2005	0.113	0.079	0.058	0.036
	(0.019)***	(0.016)***	(0.015)***	(0.017)**
Baseline log employment	-0.275	-0.328	-0.477	-0.496
	(0.009)***	(0.008)***	(0.013)***	(0.014)***
Population	0.000	0.000	0.000	0.000
	(0.000)***	(0.000)**	(0.000)**	(0.000)***
Share of land irrigated			0.099	0.078
			(0.024)***	(0.026)***
Log(land area)			0.141	0.126
			(0.008)***	(0.008)***
Distance from town			-0.002	-0.002
			(0.000)***	(0.000)***
Baseline number of industries			0.024	0.026
			(0.002)***	(0.002)***
Constant	1.115	1.734	1.305	1.377
	(0.030)***	(0.063)***	(0.078)***	(0.078)***
N	48216	48216	46720	34888
r2	0.13	0.17	0.21	0.22

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. This table presents OLS estimates of the relationship between log employment growth (1998-2005) and treatment, as defined as having received a completed PMGSY road by 2005. The sample is all locations that received a PMGSY road before 2012. Column 1 presents the estimate only controlling for 1998 (log) employment and village population. Column 2 introduces state fixed effects. Column 3 introduces standard village level controls of share of land irrigated, log land area, distance from nearest town and number of non-farm industries present in 1998. Column 4 limits to villages in which the largest habitation had fewer than 1500 people. Standard errors are clustered at the district level.

nearest town and number of non-farm industries present in 1998. Column 4 limits to villages in which the largest habitation had fewer than 1500 people. Standard errors are clustered at the district level.

The table shows that villages that received a new PMGSY road by 2005 on average had non-farm employment growth that was 3-12 log points higher than villages that did not receive a new road. The point estimates fall as more controls are included, suggesting that selection of villages for roads is non-random. In particular, the falling coefficient as state fixed effects are included suggests that higher growth states were more likely to implement the program early - this is consistent with reports that state administrative capacity played an important role in early implementation of PMGSY.

Table 3.4: *First stage: RD estimates of effect of population threshold on probability of new road*

	(1)	(2)	(3)	(4)
1(Pop $\geq$ 500)	0.034	0.031	0.030	0.034
	(0.010)***	(0.010)***	(0.010)***	(0.016)**
Population	-0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Population * 1(Pop $\geq$ 500)	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.001)
Share of land irrigated			-0.004	-0.004
			(0.007)	(0.007)
Log(land area)			0.001	0.001
			(0.004)	(0.004)
Distance from town			-0.000	-0.000
			(0.000)	(0.000)
Baseline number of industries			0.001	0.001
			(0.000)*	(0.000)*
Constant	0.016	0.012	0.010	0.007
	(0.006)***	(0.006)**	(0.018)	(0.018)
N	28747	28747	26440	26440
r2	0.01	0.02	0.02	0.02

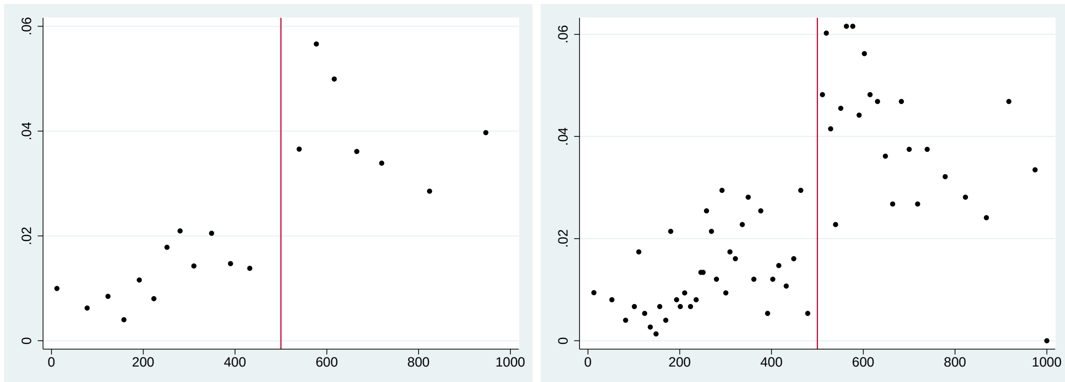
$^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$. The table presents regression discontinuity estimates from Equation 3.3 of the effect of PMGSY prioritization on a village's likelihood of receiving a road. The dependent variable is an indicator variable that is set to 1 if a village received a road by 2005. The running variable is the population of the largest habitation in the village, and the treatment variable is an indicator that is set to 1 if the largest habitation has a population greater than or equal to 500. Columns 1-3 show estimates from the local linear specification, and column 4 includes a quartic in population. Column 2 adds state fixed effects, and column 3 adds village level controls from the 1991 population census, which are log land area, share of land that is irrigated, distance to the nearest town, and the number of different industries in the village in 1998. The value 500 has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the treatment effect. Standard errors are clustered at the district level.

3.6.2 Regression discontinuity

Table 3.4 presents regression discontinuity estimates of the effect of PMGSY prioritization on a village's likelihood of receiving a road. Habitations with a population greater than 500 should have received higher priority than habitations with population under 500. The dependent variable in these regressions is an indicator variable that is set to 1 if a village received a road by 2005. The running variable is the population of the largest habitation in the village. The treatment variable is an indicator that is set to 1 if the largest habitation has a population greater than or equal to 500. The value 500 has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the treatment effect.

Columns 1-3 show estimates from the local linear specification (Equation 3.3). Column 1 is the

Figure 3.1: *First stage: population threshold and new rural roads*



The figure shows the share of habitations that received a road, by population. Each point represents approximately 1000 habitations in the top panel, and 300 habitations in the bottom panel. The PMGSY instructed states to target roads to habitations with population greater than 500, the value indicated by the solid vertical line.

baseline specification without controls. Column 2 adds state fixed effects, and column 3 adds village level controls from the 1991 population census, which are log land area, share of land that is irrigated, distance to the nearest town, and the number of different industries in the village in 1998. Column 4 adds a 4th degree polynomial in habitation population. The estimations show a robust and highly significant effect of the population threshold on the probability of a village receiving a new road. A village is on average 3% more likely to receive a new road if the population of the largest habitation is just above 500 than if the population is just below 500. Figure 3.1 depicts graphically the increase in probability of receiving a new road when the largest habitation is just above the threshold.

The bandwidth used in all specifications is 250, so the sample for the estimation are villages with a largest habitation in the range of 250-750, but results are robust to alternate bandwidth choices. All standard errors are clustered at the district level to account for spatial correlation. Controls and fixed effects are not necessary for identification, but their inclusion increases the efficiency of the estimator.

Table 3.5 presents instrumental variable regression discontinuity estimates of the effect of receiving a new road under the PMGSY on log non-farm employment growth from 1998-2005. As above, the running variable is the population of the largest habitation in the village, and a dummy variable indicating population greater than 500 instruments for a village's receiving a PMGSY road by 2005. Columns 1-3 show local linear specifications. Column 2 adds state fixed effects, column 3 adds village-level controls, and column 4 adds a quartic in population. The estimates show a very large and statistically significant effect of a new feeder road on rural non-farm employment. A new

Table 3.5: *RD estimates of effect of new road on employment growth*

	(1)	(2)	(3)	(4)
r05	2.355	1.893	1.282	1.351
	(0.528)***	(0.530)***	(0.502)**	(0.678)**
Population	-0.000	-0.000	-0.000	-0.000
	(0.000)**	(0.000)*	(0.000)***	(0.000)***
Population * 1(Pop $\geq$ 500)	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Share of land irrigated			0.193	0.192
			(0.021)***	(0.022)***
Log(land area)			0.216	0.215
			(0.008)***	(0.009)***
Distance from town			-0.003	-0.003
			(0.000)***	(0.000)***
Baseline number of industries			0.013	0.013
			(0.001)***	(0.002)***
Constant	1.224	1.520	0.643	0.644
	(0.022)***	(0.046)***	(0.064)***	(0.077)***
N	19288	19288	18972	16196
r2	0.00	0.05	0.17	0.17

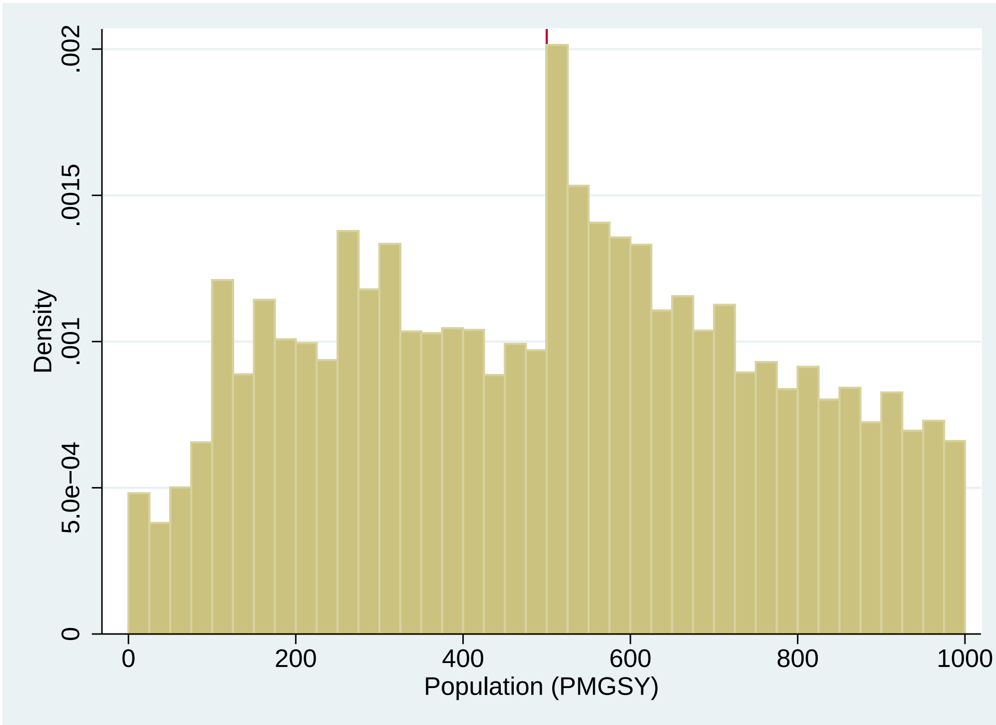
$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. The table presents instrumental variable regression discontinuity estimates of the effect of receiving a new road under the PMGSY on log non-farm employment growth from 1998-2005. The dependent variable is village level non-farm employment growth from 1998-2005. The running variable is the population of the largest habitation in the village, and the treatment indicator of population greater than 500 instruments for a village receiving a PMGSY road by 2005. Columns 1-3 show estimates from the local linear specification, and column 4 includes a quartic in population. Column 2 adds state fixed effects, and column 3 adds village level controls from the 1991 population census, which are log land area, share of land that is irrigated, distance to the nearest town, and the number of different industries in the village in 1998. The value 500 has been subtracted from population values, so that the coefficient on the uninteracted treatment variable is the estimate of the treatment effect. All columns include a control for log non-farm employment in 1998. Standard errors are clustered at the district level.

feeder road more than doubles the growth of non-farm employment. Effect sizes range from 120 to 240 log points, reflecting the high variation in growth rates across villages.[8] Controls, clustering and bandwidth are the same as in Table 3.4. All regressions include an additional control for non-farm employment in 1998.

A standard regression discontinuity validity test is that the density of the running variable is continuous across the treatment threshold. The standard test recommended by McCrary (2008) cannot be applied, because the population threshold affects both the probability of receiving a new

[8]We omit villages with non-farm employment less than 10 or greater than 200 in 1998. The median village has employment of 180, so this effect is not driven by growth in villages with very small levels of employment.

Figure 3.2: *Histogram of PMGSY-project population (SP)*

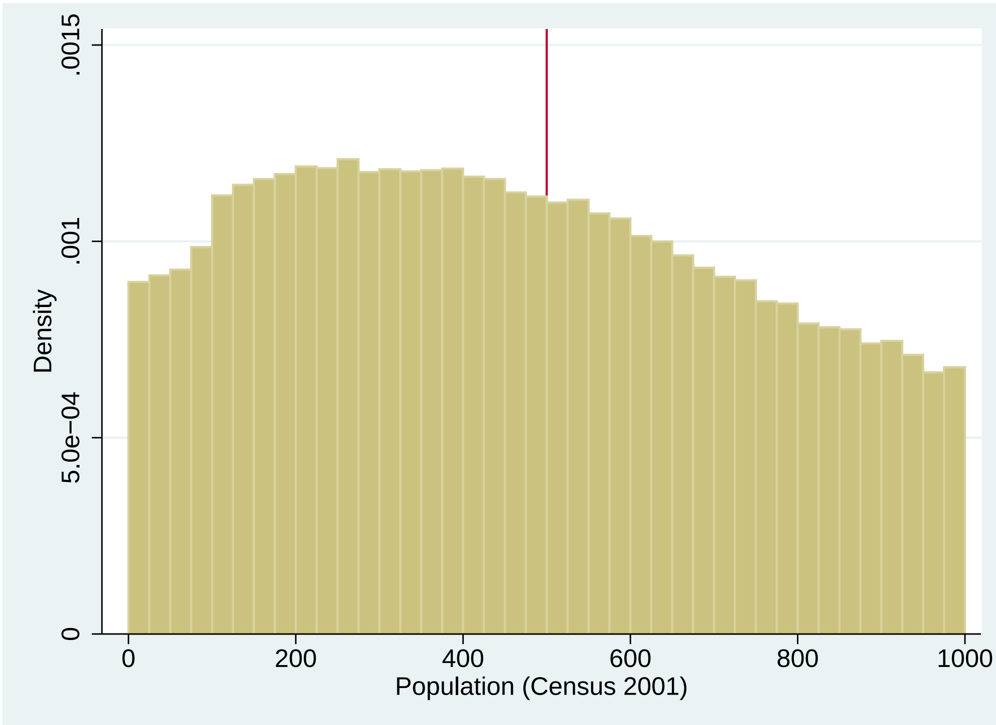


The figure displays a histogram of habitation population as recorded in the PMGSY project data. The density bulges at 250 and 500 reflect the project's prioritization of villages with population greater than 500, and then greater than 250.

road, and selection into the RD sample.[9] The histogram of PMGSY-listed populations of habitations that received roads by 2011 (the RD sample) is displayed in Figure 3.2, with clear discontinuous increases in density at the population thresholds of 250 and 500. Figure 3.3 shows the histogram of village population in the 2001 population census, where there is no evident sign of a discontinuity at either of the PMGSY thresholds.

[9]The sample for the RD is all villages in which a PMGSY project was sanctioned by 2012. Habitations with populations greater than 500 are more likely to be prioritized and receive a road by 2005, which drives our empirical result. However, habitations with population greater than 500 are also more likely to receive a road by 2011, which drives selection into the sample. An alternate approach which would not face this issue would be to use the entire set of Indian villages as a control group - but with fewer than 40,000 out of 600,000+ villages connected, the increased probability of receiving a road under PMGSY is too small to detect.

Figure 3.3: *Histogram of population (population census 2001)*



The figure displays a histogram of village population as recorded in the 2001 population census. The PMGSY eligibility thresholds of 250 and 500 do not show obvious changes in density.

3.6.3 IV

In this section we present results from the IV specification described in Section 3.4.3. We first verify that within-district village population rank appears to be a valid instrument and then discuss results of this estimation strategy.

We define our sample to be villages listed as not having paved roads at the time of the 2001 Population Census. This is for two reasons. First, villages unconnected by paved all-weather roads to the road network are listed as the highest priority under the PMGSY. Second, we can be confident that roads to such villages are in fact new roads and not upgrades of existing roads, in order to be able to more cleanly interpret the results. The sample is further restricted to villages in districts in which the PMGSY built roads to more than 5% of villages by 2005. Finally, because few villages with very high within-district population ranks (very small villages) receive roads under the PMGSY, we drop villages with population ranks above 300. These sample definitions yield a sample of 12073 eligible villages in our rank IV sample, out of which 1910 received PMGSY roads by 2005.

First stage and reduced form

We define our rank instrument two ways. The first is a simple field rank that assigns a village a rank of the number of villages in the sample with populations greater than the village, plus one. The second creates a binary variable out of this rank that takes on the value 1 when a village rank is less than 75. Table 3.6 presents the results of the first stage, regressing an indicator for the completion of a PMGSY road by 2005 on these two rank instruments. For simplicity of presentation, the rank variable has been divided by 100. Columns 1 and 2 present the effect of rank on the probability that a village receives a road by 2005 with quadratic and quartic polynomial population controls, respectively. A reduction of rank of 100 is associated with a 4 to 5 percentage point increase in the probability of receiving a road, controlling for population. Columns 3 and 4 present the effect of our binary instrument on the likelihood of receiving a PMGSY road by 2005: being in the top 75 villages within one's district is, after controlling for population and other village characteristics, associated with an approximately 5.5 percentage point increase in the probability of receiving a PMGSY road. The results are little changed by adding additional population controls, which is reassuring for the validity of the instruments.

Table 3.7 presents the reduced form results. The dependent variable in all four columns is log nonfarm employment growth. Columns 1 and 2 show that an increase in rank of 100 (decrease in prioritization for PMGSY road) is associated with a reduction of approximately 7 log points of employment growth, with negligible change when going from quadratic to quartic population controls. Likewise, columns 3 and 4 show that having a rank of less than 75 is associated with an increase of approximately 6 log points of employment growth.

One concern is that the rank instrument may be correlated with prioritization not only in the PMGSY but in other government programs that are also carried out at the district level. We thus define a placebo sample of villages with equivalent rankings as our rank IV sample but in districts that did not construct PMGSY roads in more than 1% of villages. Table 3.8 presents the results of the same reduced form specification discussed above but in this placebo sample. We find insignificant effects of both the continuous and binary population rank instruments on village nonfarm employment growth. We take this as suggestive, if not conclusive, evidence that the village population rank affects employment growth only through its impact on the likelihood of receiving a PMGSY road.

Table 3.6: *First stage effect of rank on probability of receiving road*

	1	2	3	4
Pop rank (no road)	-0.047	-0.043		
	(0.007)***	(0.007)***		
Top pop (no road)			0.056	0.055
			(0.008)***	(0.008)***
Ln baseline employment	-0.000	-0.000	0.001	0.000
	(0.005)	(0.005)	(0.005)	(0.005)
Village pop	0.000	0.002	0.000	0.002
	(0.000)***	(0.001)***	(0.000)***	(0.001)***
Village pop ^{2}	-0.001	-0.016	-0.001	-0.017
	(0.000)***	(0.006)***	(0.000)***	(0.006)***
Village pop ^{3}		0.053		0.058
		(0.023)**		(0.023)**
Village pop ^{4}		-0.006		-0.007
		(0.003)**		(0.003)**
Pop growth 1991-2001	-0.065	-0.066	-0.066	-0.067
	(0.013)***	(0.013)***	(0.013)***	(0.013)***
Irrigation share	0.018	0.017	0.015	0.013
	(0.012)	(0.012)	(0.012)	(0.012)
Ln land area	0.009	0.008	0.009	0.008
	(0.004)**	(0.004)*	(0.004)**	(0.004)*
Distance from town	-0.001	-0.001	-0.001	-0.001
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Diversity (1998)	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
N	12073	12073	12073	12073
r2	0.11	0.11	0.11	0.11

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents first stage regressions using within-district population rank as a predictor of receipt of a PMGSY road by 2005. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 3.7: *Reduced form effect of rank on employment growth*

	1	2	3	4
Pop rank (no road)	-0.073	-0.069		
	(0.016)***	(0.016)***		
Top pop (no road)			0.060	0.059
			(0.018)***	(0.018)***
Ln baseline employment	-0.610	-0.610	-0.608	-0.608
	(0.011)***	(0.011)***	(0.011)***	(0.011)***
Village pop	0.001	0.003	0.001	0.003
	(0.000)***	(0.001)**	(0.000)***	(0.001)***
Village pop ^{2}	-0.001	-0.024	-0.001	-0.026
	(0.000)**	(0.013)*	(0.000)***	(0.013)**
Village pop ^{3}		0.087		0.094
		(0.050)*		(0.049)*
Village pop ^{4}		-0.011		-0.012
		(0.007)		(0.007)*
Pop growth 1991-2001	-0.122	-0.123	-0.123	-0.124
	(0.034)***	(0.034)***	(0.034)***	(0.034)***
Irrigation share	0.107	0.105	0.104	0.102
	(0.026)***	(0.026)***	(0.026)***	(0.026)***
Ln land area	-0.040	-0.041	-0.041	-0.042
	(0.010)***	(0.010)***	(0.010)***	(0.010)***
Distance from town	-0.002	-0.002	-0.002	-0.002
	(0.000)***	(0.000)***	(0.000)***	(0.000)***
Diversity (1998)	0.011	0.011	0.011	0.011
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
N	12073	12073	12073	12073
r2	0.33	0.33	0.33	0.33

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. This table presents reduced form regressions using within-district population rank as a predictor of log employment growth between 1998 and 2005. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 3.8: *Reduced form effect of rank on employment growth (placebo sample)*

	1	2	3	4
Pop rank (no road)	0.019	0.019		
	(0.033)	(0.033)		
Top pop (no road)			0.023	0.022
			(0.045)	(0.045)
Ln baseline employment	-0.585	-0.585	-0.588	-0.587
	(0.024)***	(0.024)***	(0.024)***	(0.024)***
Village pop	0.001	0.002	0.001	0.002
	(0.000)***	(0.003)	(0.000)***	(0.003)
Village pop ^{2}	-0.001	-0.013	-0.001	-0.012
	(0.000)***	(0.026)	(0.000)***	(0.026)
Village pop ^{3}		0.041		0.038
		(0.099)		(0.099)
Village pop ^{4}		-0.005		-0.005
		(0.014)		(0.014)
Pop growth 1991-2001	-0.059	-0.058	-0.058	-0.057
	(0.067)	(0.067)	(0.067)	(0.067)
Irrigation share	-0.130	-0.131	-0.133	-0.133
	(0.055)**	(0.055)**	(0.055)**	(0.055)**
Ln land area	-0.023	-0.023	-0.023	-0.023
	(0.020)	(0.020)	(0.020)	(0.020)
Distance from town	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Diversity (1998)	0.000	0.000	0.000	0.000
	(0.005)	(0.005)	(0.005)	(0.005)
N	2871	2871	2871	2871
r2	0.35	0.35	0.35	0.35

$^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. This table presents reduced form regressions using within-district population rank as a predictor of log employment growth between 1998 and 2005 for a placebo sample of villages in districts not receiving many PMGSY roads. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in less than 1% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Rank IV estimates

Table 3.10 presents the results of the IV estimation based on Equation 3.4. As in the preceding tables, the first two columns use $RANK$ as the excluded instrument for the construction of a road, while columns 3 and 4 use the binary low rank variable. We find that a new road leads to highly significant positive increases in nonfarm employment growth. Regardless of specification, a new road is associated with an increase of employment growth of over 100 log points. Table 3.9 seeks to unpackage these results, instrumenting for new road with only the binary rank variable. Column 1 estimates the effect of a new road on the number of economic establishments (in logs); in contrast to the highly significant positive results for employment growth, the estimated effect on firm count is smaller and less significant. Consistent with this result, column 2 finds that a new road is associated with a large and significant increase in firm size: the mean firm employs more than 1 additional employee than in villages that do not receive roads. Column 3 estimates the treatment effect on the percent change in economic diversity of the village. We follow Hidalgo and Hausmann (2009) in defining the diversity of a village economy as the number of industries present in the village.[10] The outcome variable is defined as the proportional change in the economic diversity of a village, top-coded at 4 (400% increase) to reduce the effect of outliers. We find that a new road is associated with a positive significant increase of approximately 0.65 industries per village. Column 4 estimates the effect of a new road on the village share of employment in informal firms, defined as those not registered with the Indian government. We find a large increase in informality, although it is unclear how best to interpret this result. One hypothesis is that high transportation costs keep firms inefficiently small and informal by denying them access to larger markets. This theory would predict increased formalization in response to the construction of a new road, but if the employment response to falling transport costs is faster than the process of formalization, we should expect to see a short-term rise informality. Future work with the Economic Census of 2012 should enable us to test this hypothesis.

Table 3.11 considers heterogeneity of our estimated treatment effect by village characteristics. Theoretically, we expect that firms with better access productive inputs to grow most in response to a fall in transportation costs. Column 1 presents the full sample estimate of the effect of a road on total

[10]The 1998 Economic Census reports 3 digit NIC-1987 codes for each establishment, while the 2005 Economic Census uses 4-digit NIC-2004 codes. We use correspondence tables published by the Ministry of Statistics and Programme Implementation to generate a list of 217 self-contained industries, which we use as the basis of all industry-level analysis in this paper.

Table 3.9: *IV effect of road on firm characteristics*

	Firm Count	Firm Size	Diversity	Informality
New road	0.510	1.423	0.654	0.313
	(0.316)	(0.422)***	(0.307)**	(0.110)***
Ln baseline employment				
Village pop	0.002	-0.002	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)**
Village pop ^{2}	-0.012	0.016	0.009	0.008
	(0.013)	(0.016)	(0.013)	(0.004)*
Village pop ^{3}	0.046	-0.056	-0.030	-0.027
	(0.051)	(0.062)	(0.050)	(0.017)
Village pop ^{4}	-0.007	0.007	0.004	0.003
	(0.007)	(0.008)	(0.007)	(0.002)
Pop growth 1991-2001	-0.109	0.111	-0.043	0.008
	(0.039)***	(0.048)**	(0.039)	(0.013)
Irrigation share	0.095	-0.055	0.137	-0.003
	(0.025)***	(0.032)*	(0.025)***	(0.008)
Ln land area	-0.025	-0.066	-0.026	0.002
	(0.010)**	(0.012)***	(0.010)**	(0.003)
Distance from town	-0.001	0.000	-0.001	0.000
	(0.000)***	(0.001)	(0.000)*	(0.000)
Diversity (1998)	0.011	0.007	-0.081	0.002
	(0.002)***	(0.002)***	(0.002)***	(0.001)***
o.Diversity (1998)			0.000	
			(.)	
N	12073	11891	12073	12073
r2	0.32	.	0.18	0.01

$^{*}p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$. This table presents instrumental variable regression results for which the endogenous regressor is a dummy variable indicating the construction of a PMGSY road by 2005, instrumented by a dummary variable indicating that within-district, within-sample population rank is less than 75. Dependent variables for columns 1 through 4 are, respectively: count of economic establishments, mean firm size (number of employees), proportional change in economic diversity and 2005 share of employment in informal firms. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

Table 3.10: *IV effect of road on employment growth*

	1	2	3	4
New road	1.542	1.613	1.071	1.081
	(0.400)***	(0.446)***	(0.356)***	(0.364)***
Ln baseline employment	-0.610	-0.610	-0.609	-0.609
	(0.014)***	(0.014)***	(0.013)***	(0.013)***
Village pop	0.000	-0.000	0.000	0.001
	(0.000)	(0.002)	(0.000)**	(0.002)
Village pop ^{2}	0.000	0.002	-0.000	-0.007
	(0.000)	(0.017)	(0.000)	(0.015)
Village pop ^{3}		0.002		0.032
		(0.066)		(0.059)
Village pop ^{4}		-0.001		-0.005
		(0.009)		(0.008)
Pop growth 1991-2001	-0.021	-0.017	-0.052	-0.052
	(0.046)	(0.048)	(0.042)	(0.043)
Irrigation share	0.078	0.078	0.087	0.087
	(0.032)**	(0.032)**	(0.029)***	(0.029)***
Ln land area	-0.054	-0.054	-0.051	-0.051
	(0.013)***	(0.013)***	(0.012)***	(0.012)***
Distance from town	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)**	(0.001)**
Diversity (1998)	0.011	0.011	0.011	0.011
	(0.003)***	(0.003)***	(0.002)***	(0.002)***
N	12073	12073	12073	12073
r2	.	.	0.15	0.15

$^* p < 0.10, ^{**} p < 0.05, ^{***} p < 0.01$. This table presents instrumental variable regression results for which the dependent variable is log employment growth (1998-2005) and the endogenous regressor is a dummy variable indicating the construction of a PMGSY road by 2005, instrumented by population rank. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

employment, while columns 2 through 7 present subsample results. We find suggestive evidence that electricity appears to be an important complementary input to a village feeder road: while we estimate a very small (if noisy) estimate for villages without electrical supply, we find significant, statistically larger effects in villages with electricity supply. This result should be interpreted with some caution, given that our identification strategy provides exogenous variation in road placement, not in the interaction of road construction and availability of electricity. Next, we estimate effects for villages above and below the median distance to the nearest town, which for our sample is 20 km. We estimate a larger treatment effect for villages far from towns, although both estimates are within one standard deviation of our full-sample estimate and are not statistically different from each other. Likewise, we find no significant difference in our estimates when dividing our sample on human capital, as measured by literacy.

Given the very high costs of infrastructure projects such as roads, it is of great interest to policymakers to understand the economic impact of such projects in terms of money spent rather than per project. The OMMS described in Section 3.5 contains data on total spending per road, which we use to construct a village road expenditure variable that takes on the value 0 if a road is not built in the village by 2005 and otherwise equals the sum of all PMGSY road spending in habitations contained in that village.[11] As our objective is to estimate the cost of job creation, we define our outcome variable to be the level change in employment between 1998 and 2005, rather than log growth as in the preceding tables. Table 3.12 presents the results of this estimation. Our estimates range, depending on the instrument used, between approximately 500 and 700 nonfarm jobs per million dollars spent on road construction, which results in a per job cost of between approximately 1430 and 2000 USD. Expressed differently, the per job cost of the PMGSY is roughly two to three times India's GDP per capita in 2005 ($731.70). How should we interpret this result? As the Economic Census was conducted in late 2005 and early 2006, it is safe to assume that nearly all of the roads listed as completed by 2005 were finished by the time of data collection; thus, these results should be understood not as temporary jobs related to road construction, but rather short- to medium-term effects. These estimates also assume that the net employment effect of PMGSY spending in locations not receiving roads is zero. This is a strong assumption – one we intend to investigate in future work – as there may be either crowd-in or crowd-out of economic activity in villages and towns that are

[11] We generate total road spending village in million USD, using an exchange rate of 44.06 INR per USD, the average for 2005. In the case of roads that connect multiple habitations, spending is allocated equally between the habitations.

Table 3.11: *IV effect of road on employment growth by village characteristics*

	All	No Electricity	Electricity	Far	Near	Low Literacy	High Literacy
New road	1.081	0.019	1.728	1.293	0.846	1.415	0.559
	(0.364)***	(0.699)	(0.511)***	(0.537)**	(0.536)	(0.493)***	(0.651)
Ln baseline employment	-0.609	-0.680	-0.608	-0.628	-0.596	-0.618	-0.616
	(0.013)***	(0.024)***	(0.019)***	(0.019)***	(0.017)***	(0.018)***	(0.019)***
2001 Population	0.001	-0.004	0.005	-0.003	0.005	-0.001	0.003
	(0.002)	(0.005)	(0.002)*	(0.003)	(0.002)**	(0.003)	(0.002)
2001 Population ^{2}	-0.007	0.040	-0.042	0.027	-0.047	0.008	-0.021
	(0.015)	(0.044)	(0.022)*	(0.023)	(0.020)**	(0.022)	(0.021)
2001 Population ^{3}	0.032	-0.169	0.175	-0.109	0.194	-0.020	0.075
	(0.059)	(0.159)	(0.087)**	(0.089)	(0.078)**	(0.085)	(0.081)
2001 Population ^{4}	-0.005	0.025	-0.025	0.016	-0.028	0.002	-0.010
	(0.008)	(0.021)	(0.012)**	(0.012)	(0.011)***	(0.012)	(0.011)
Pop growth 1991-2001	-0.052	-0.068	0.021	0.040	-0.157	-0.109	0.028
	(0.043)	(0.063)	(0.060)	(0.064)	(0.056)***	(0.058)*	(0.063)
Irrigation share	0.087	0.132	0.042	0.032	0.175	0.217	-0.075
	(0.029)***	(0.050)***	(0.051)	(0.040)	(0.043)***	(0.043)***	(0.039)*
Ln land area	-0.051	-0.045	-0.046	-0.064	-0.030	-0.046	-0.036
	(0.012)***	(0.018)**	(0.019)**	(0.016)***	(0.017)*	(0.016)***	(0.022)
Distance from town	-0.001	-0.002	-0.001	-0.002	-0.001	-0.001	-0.001
	(0.001)**	(0.001)**	(0.001)	(0.002)	(0.001)*	(0.001)	(0.001)
Diversity (1998)	0.011	0.015	0.011	0.014	0.008	0.011	0.011
	(0.002)***	(0.005)***	(0.004)***	(0.003)***	(0.003)**	(0.004)***	(0.003)***
N	12073	3106	6453	6646	5427	6886	5187
r2	0.15	0.34	.	0.08	0.22	0.05	0.26

$^{*} p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. This table presents instrumental variable estimates of the effect of the construction of a PMGSY road on log employment growth (1998-2005) for different subsamples in order to estimate heterogeneous treatment effects. All subsamples are defined based on 2001 Population Census characteristics. Column 1 presents full sample results. Columns 2 and 3 are for villages with and without electricity supply, respectively. Column 4 presents results for villages with below median distance from the nearest town (20km). Column 5 presents results for villages above median distance from a town. Column 6 presents results for villages with below median literacy, and Column 7 for villages with above-median literacy. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

now better connected to PMGSY villages. Finally, all estimates assume that control villages do not receive any PMGSY road spending, an unlikely result given the multiple years that road construction generally requires, the multiple habitations per PMGSY road project and the many roads that were completed in 2006 and 2007. For this last reason, we consider our estimates to be likely lower bounds on the true effect of road construction in this context.

3.7 Conclusion

Universal access to paved roads, easily taken for granted in many rich countries, is far from a reality for many of the world's poor. High transportation costs inhibit gains from the division of labor, economies of scale and comparative advantage. Nevertheless, little is known of the economic effects of road provision on rural economic activity. Theoretically, roads could facilitate migration to urban areas and the specialization of village economies in agriculture; alternatively, lower transportation costs could cause the emergence and expansion of rural economic activities, with potentially large consequences for economic development, urbanization and the spatial distribution of economic activity.

In this paper we estimate the economic impacts of a large-scale program that seeks to provide near universal access to paved "all-weather" roads in rural India. The program design provides two sources of exogenous variation to allow us to overcome the usual challenge of endogeneity of large infrastructure projects. First, the program calls for highest priority to be given to habitations above population thresholds, which may be 250, 500 or 1000 depending on the area. This creates a discontinuity in the probability of receiving a road at these cutoffs, allowing us to use a fuzzy regression discontinuity design to estimate the impact of these roads. A second identification strategy takes advantage of the fact that habitations are prioritized not only by population but also by the relative population rank within a district: a village of a certain size is more likely to receive a road if there are fewer larger villages within its district than an equivalent village that has a lower population rank. We instrument for road construction using this rank, conditioning on population, to provide a second set of estimates of the impact of village feeder roads.

We find that the provision of a new, paved village approach road produces significantly faster nonfarm employment growth. Rather than a simple proportional expansion of existing economic activity, new industries form in villages that receive roads. Firm size also grows, which we take as supporting evidence of the theory that in the presence of high transportation costs, firms are

Table 3.12: *IV effect of road spending on level change in employment*

	1	2	3	4
Road cost (million USD)	481.839	479.030	709.091	715.636
	(303.054)	(324.152)	(333.751)**	(341.696)**
Employment (EC98)	-0.356	-0.356	-0.356	-0.356
	(0.029)***	(0.029)***	(0.030)***	(0.030)***
Village pop	0.004	0.042	-0.002	0.000
	(0.012)	(0.145)	(0.013)	(0.148)
Village pop ^{2}	0.059	-0.300	0.076	0.001
	(0.034)*	(1.358)	(0.035)**	(1.382)
Village pop ^{3}		1.422		0.486
		(5.428)		(5.516)
Village pop ^{4}		-0.199		-0.092
		(0.778)		(0.789)
Pop growth 1991-2001	-1.652	-1.679	-0.373	-0.339
	(2.850)	(2.925)	(3.095)	(3.134)
Irrigation share	4.674	4.661	4.815	4.845
	(2.336)**	(2.341)**	(2.429)**	(2.439)**
Ln land area	-5.673	-5.673	-6.180	-6.180
	(1.123)***	(1.138)***	(1.206)***	(1.201)***
Distance from town	-0.093	-0.093	-0.097	-0.098
	(0.042)**	(0.042)**	(0.044)**	(0.044)**
Diversity (1998)	-0.018	-0.019	0.000	0.002
	(0.247)	(0.248)	(0.250)	(0.251)
N	11833	11833	11833	11833
r2	0.08	0.08	.	.

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. This table presents instrumental variable regression results for which the dependent variable is log employment growth (1998-2005) and the endogenous regressor is the amount of spending on a PMGSY road by 2005. For roads are not completed by 2005, this number is 0. All regressions weight by baseline log employment and include state fixed effects, as well as the following village level controls: 1991-2001 population growth, share of land irrigated, log land area, distance from nearest town and economic diversity as measured in the 1998 Economic Census. Columns 1 and 2 define population rank as the count of in-district, in-sample villages with greater populations, plus 1. Columns 3 and 4 use a binary variable that takes the value 1 when this rank is less than 75. Columns 1 and 3 use quadratic population controls, while Columns 2 and 4 use quartic population controls. Sample is comprised of villages listed as not having a paved approach road in the 2001 Population Census. We further limit to villages in districts that constructed PMGSY roads in more than 5% of villages and villages whose within-district population rank is less than 300. Reported standard errors are heteroskedasticity-robust.

inefficiently small. The responsiveness of nonfarm employment growth to lower transportation costs appears to be strongest in villages that also have access to the supply of electricity, suggesting significant complementarities between these different infrastructural investments. Finally, we provide some of the first well-identified estimates of the cost effectiveness of rural road construction: one job is created for every $1400 to $2000 in road construction costs, suggesting very high returns to such investments.

Future work will allow us to further disentangle the channels by which rural roads promote village nonfarm employment. The 2011 Population Census, not yet available at the village level, will provide the data necessary to examine the impact of roads on migration and agricultural employment, neither of which is covered in the Economic Census. We intend to use the 2012 Economic Census, the collection of which is still in progress, to differentiate between the short- to medium-run effects presented in this paper and sustained, longer-run changes to village economic activity.

References

Aghion, P, L Boustan, C Hoxby, and J Vandenbussche, "The Causal Impact of Education on Economic Growth: Evidence from U.S.," 2009.

Albouy, David, "Partisan Representation in Congress and the Geographic Distribution of Federal Funds," 2009.

Alexeev, Michael and Robert Conrad, "The elusive curse of oil," *The Review of Economics and Statistics*, 2009, *91* (3), 586–598.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "Students today, teachers tomorrow: Identifying constraints on the provision of education," *Journal of Public Economics*, April 2013, *100*, 1–14.

Angrist, Joshua D. and Victor Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, 1999, *114* (2).

Ansolabehere, Stephen and James M Snyder, "Party Control of State Government and the Distribution of Public Expenditures," *Scandinavian Journal of Economics*, December 2006, *108* (4), 547–569.

Aragon, F M and J P Rud, "Natural Resources and Local Communities: Evidence from a Peruvian Gold Mine," *American Economic Journal: Economic Policy*, 2013, *5* (2).

Arulampalam, Wiji, Sugato Dasgupta, Amrita Dhillon, and Bhaskar Dutta, "Electoral goals and center-state transfers: A theoretical model and empirical evidence from India," *Journal of Development Economics*, January 2009, *88* (1), 103–119.

Asher, Sam and Paul Novosad, "Dirty Politics: Natural resource wealth and politics in India," 2013.

Atkin, David and Dave Donaldson, "Who's Getting Globalized? The Size and Nature of Intranational Trade Costs," 2012.

Banerjee, Abhijit, "A theory of misgovernance," *The Quarterly Journal of Economics*, 1997, *112* (4).

____ **and Rohini Somanathan**, "The political economy of public goods: Some evidence from India," *Journal of Development Economics*, March 2007, *82* (2), 287–314.

____ **, Esther Duflo, and Nancy Qian**, "On the Road: Access to Transportation Infrastructure and Economic Growth in China," 2012.

____ **, Lakshmi Iyer, and Rohini Somanathan**, "History, Social Divisions, and Public Goods in Rural India," *Journal of the European Economic Association*, 2005, *3* (2), 639–647.

____ **, Santosh Kumar, and Rohini Pande**, "Connectivity and Rural Development: Examining India's Rural Road Building Scheme," 2012.

Besley, Timothy and Robin Burgess, "Can Labor Regulation Hinder Economic Performance? Evidence from India," *Quarterly Journal of Economics*, February 2004, *119* (1), 91–134.

——, **Rohini Pande, and Vijayendra Rao**, "Just Rewards? Local Politics and Public Resource Allocation in South India," *The World Bank Economic Review*, October 2011, *26* (2), 191–216.

——, ——, **Lupin Rahman, and Vijayendra Rao**, "The Politics of Public Good Provision: Evidence from Indian Local Governments," *Journal of the European Economic Association*, 2004, *2* (2/3).

Black, Dan, Terra McKinnish, and Seth Sanders, "The Economic Impact of the Coal Boom and Bust," *The Economic Journal*, April 2005, *115* (503), 449–476.

Brollo, Fernanda and Tommaso Nannicini, "Tying Your Enemy's Hands in Close Races: The Politics of Federal Transfers in Brazil," *American Political Science Review*, 2012.

Carrington, W.J., "The Alaskan labor market during the pipeline era," *Journal of Political Economy*, 1996, *104* (1), 186–218.

Carvalho, D R, "The Real Effects of Government-Owned Banks: Evidence from an Emerging Market," 2010.

Casaburi, Lorenzo, Rachel Glennerster, and Tavneet Suri, "Rural Roads and Intermediated Trade: Regression Discontinuity Evidence from Sierra Leone," 2012.

Caselli, Francesco and Guy Michaels, "Do oil windfalls improve living standards? Evidence from Brazil," *American Economic Journal: Applied Economics*, 2013, *5* (1), 208–238.

Chaudury, Shoma, "Trust Votes And Knife Play," *Tehelka Magazine*, May 2009.

Chhibber, Pradeep, Sandeep Shastri, and Richard Sisson, "Federal Arrangements and the Provision of Public Goods in India," *Asian Survey*, June 2004, *44* (3), 339–352.

Cohen, Lauren, Joshua Coval, and Christopher Malloy, "Do Powerful Politicians Cause Corporate Downsizing," *Journal of Political Economy*, 2011, *119* (6), 1015–1060.

Cole, Shawn, "Fixing Market Failures or Fixing Elections? Agricultural Credit in India," *American Economic Journal: Applied Economics*, January 2009, *1* (1), 219–250.

Corden, W.M., "Booming Sector and Dutch Disease Economics : Survey and Consolidation," *Oxford Economic Papers*, 2012, *36* (3), 359–380.

Dixit, Avinash and John Londregan, "The Determinants of Success of Special Interests in Redistributive Politics," *The Journal of Politics*, 1996, *58* (4), 1132–1155.

Djankov, Simeon, Caroline Freund, and Cong S. Pham, "Trading on time," *The Review of Economics and Statistics*, 2010, *92* (1), 166–173.

Domenech, Jordi, "Mineral resource abundance and regional growth in Spain, 1860–2000," *Journal of International Development*, 2008, *20*, 1122–1135.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *October*, 2012, (October).

Ellison, Glenn and Edward L Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 1997, *105* (5), 889–927.

Fafchamps, Marcel and Forhad Shilpi, "Cities and Specialisation: Evidence from South Asia," *The Economic Journal*, 2005, *115* (503), 477–504.

Fan, Shenggen and Peter Hazell, "Returns to Public Investments in the Less-Favored Areas of India and China," *American Journal of Agricultural Economics*, 2001, *83* (5), 1217–1222.

Ferraz, Claudio and Joana Monteiro, "Does Oil Make Leaders Unaccountable? Evidence from Brazil's offshore oil boom," 2010.

Finan, Frederico S, "Political Patronage and Local Development : A Brazilian Case Study," 2004.

Fisman, Raymond, "Estimating the Value of Political Connections," *The American Economic Review*, February 2001, *91* (4), 867–8.

Foster, Andrew D and Mark R Rosenzweig, "Agricultural Productivity Growth, Rural Economic Diversity and Economic Reforms, India 1970-2000," *Economic Development and Cultural Change*, 2004.

Geological Survey of India, *Mineral Atlas of India*, Kolkata: Geological Survey of India, 2001.

Gollin, Douglas and Richard Rogerson, "Agriculture, Roads and Economic Development in Uganda," 2010.

Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano, "Policy Responses to Fiscal Restraints: A Difference-in-Discontinuities Design," 2012.

Grimmer, Justin, Eitan Hersh, Brian Feinstein, and Daniel Carpenter, "Are Close Elections Random?," 2012.

Hidalgo, César and Ricardo Hausmann, "The building blocks of economic complexity.," *Proceedings of the National Academy of Sciences of the United States of America*, June 2009, *106* (26), 10570–5.

Hoover, Gary A and Paul Pecorino, "The Political Determinants of Federal Expenditure at the State Level," *Public Choice*, 2005, *123*, 95–113.

Hornbeck, Richard and Dave Donaldson, "Railroads and American Economic Growth: A "Market Access" Approach," 2012.

Imbens, Guido and K. Kalyanaraman, "Optimal bandwidth choice for the regression discontinuity estimator," Technical Report February, National Bureau of Economic Research 2009.

Imbens, Guido W. and Thomas Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, February 2008, *142* (2), 615–635.

Indian Bureau of Mines, *Indian Minerals Year Book 2010*, Nagpur, India: IBM Press, 2011.

Iyer, Lakshmi and Anandi Mani, "Traveling Agents: Political Change and Bureaucratic Turnover in India," *The Review of Economics and Statistics*, 2012, *94* (3), 723–739.

Jayachandran, Seema, "The Jeffords Effect," *Journal of Law and Economics*, 2006, *49* (2).

Jha, Raghbendra, Sambit Bhattacharyya, Raghav Gaiha, and Shylashri Shankar, ""Capture" of anti-poverty programs: An analysis of the National Rural Employment Guarantee Program in India," *Journal of Asian Economics*, September 2009, *20* (4), 456–464.

Kasara, Kimuli, "Tax Me If You Can: Ethnic Geography, Democracy, and the Taxation of Agriculture in Africa," *American Political Science Review*, 2007, *101* (1).

Kelly, Thomas D. and Grecia R. Matos, "Historical Statistics for Mineral and Material Commodities in the United States," Technical Report, US Geological Survey 2013.

Khandker, Shaidur R. and Gayatri B Koolwal, "Estimating the Long-term Impacts of Rural Roads A Dynamic Panel Approach," 2011.

——, **Zaid Bakht, and Gayatri B. Koolwal**, "The Poverty Impact of Rural Roads: Evidence from Bangladesh," *Economic Development and Cultural Change*, 2009, *57* (4), 685–722.

Khemani, S., "Does delegation of fiscal policy to an independent agency make a difference? Evidence from intergovernmental transfers in India," *Journal of Development Economics*, 2007, *82* (2), 464–484.

Khwaja, Asim and Atif Mian, "Do lenders favor politically connected firms? Rent provision in an emerging market," *Quarterly Journal of Economics*, 2005, *120* (4), 1371–1411.

Lee, D, "Randomized experiments from non-random selection in U.S. House elections," *Journal of Econometrics*, February 2008, *142* (2), 675–697.

Lee, David and Thomas Lemieux, "Regression discontinuity designs in economics," *Journal of Economic Literature*, 2010, *48* (June), 281–355.

Limao, Nuno and Anthony J Venables, "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade," *The World Bank Economic Review*, 2001, *15* (3), 451–479.

Malatesta, Paul H and Rex Thompson, "Partially Anticipated Events: A Model of Stock Price Reactions with an Application to Corporate Acquisitions," *Journal of Financial Economics*, 1985, *14*, 237–250.

Marshall, Alfred, *Principles of Economics*, London: MacMillan, 1920.

McCrary, Justin, "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, February 2008, *142* (2), 698–714.

McDonald, Hamish, *Mahabharata in Polyester*, Sydney: University of New South Wales Press Ltd, 2010.

Mehlum, Halvor, Karl Moene, and Ragnar Torvik, "Institutions and the Resource Curse," *The Economic Journal*, 2006, *116* (January), 1–20.

Michaels, Guy, "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *The Review of Economics and Statistics*, 2008, *90* (November), 683–701.

____ , "The long term consequences of resource-based specialization," *The Economic Journal*, 2010, *121* (March), 31–57.

Ministry of Rural Development, "Working Group on Rural Roads in the 12th Five Year Plan," Technical Report 2012.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny, "The allocation of talent: Implications for growth," *Quarterly Journal of Economics*, 1991, *106* (2), 503–530.

National Rural Roads Development Agency, "Pradhan Mantri Gram Sadak Yojana - Operations Manual," Technical Report, Ministry of Rural Development, Government of India 2005.

Panagariya, Arvind, *India: The Emerging Giant*, USA: Oxford University Press, 2008.

Pritchett, Lant, "A Review of Edward Luce's "In Spite of the Gods: The Strange Rise of Modern India"," *Journal of Economic Literature*, 2009, *47* (3), 771–780.

Raballand, Gael, Rebecca Thornton, Dean Yang, Jessica Goldberg, Niall Keleher, and Annika Muller, "Are Rural Road Investments Alone Sufficient to Generate Transport Flows? Lessons from a Randomized Experiment in Rural Malawi and Policy Implications," 2011.

Ramey, Valerie A, "Can Government Purchases Stimulate the Economy," *Journal of Economic Literature*, 2011, *49* (3), 673–685.

Redding, By Stephen J and Daniel M Sturm, "The Costs of Remoteness : Evidence from German Division and Reunification," *American Economic Review*, 2008, (1999), 1766–1797.

Roberts, Brian E, "A Dead Senator Tells No Lies: Seniority and the Distribution of Federal Benefits," *American Journal of Political Science*, 1990, *34* (1), 31–58.

Robinson, James, Ragnar Torvik, and Thierry Verdier, "Political foundations of the resource curse," *Journal of Development Economics*, April 2006, *79* (2), 447–468.

Sachs, J.D. and A.M. Warner, "The curse of natural resources," *European economic review*, 2001, *45* (4-6), 827–838.

Sachs, Jeffrey and A.M. Warner, "Natural Resource Abundance and Economic Growth," in G Meier and J Rauch, eds., *Leading Issues in Economic Development*, Vol. 3, New York: Oxford University Press, 1995.

Shleifer, Andrei and Robert W. Vishny, "Politicians and Firms," *Quarterly Journal of Economics*, 1994, *109* (4), 995–1025.

Shoag, Daniel, "The Impact of Government Spending Shocks: Evidence on the Multiplier from State Pension Plan Returns," 2011.

Singh, K. P., *Whatever the odds: the incredible story of DLF*, New Delhi: HarperCollins Publishers India, 2011.

Storeygard, Adam, "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," 2012.

Sukhtankar, Sandip, "Sweetening the Deal? Political Connections and Sugar Mills in India," *American Economic Journal: Applied Economics*, 2012, *4* (3), 43–63.

The Hindu, "Khemka refutes Haryana government claims, says transfer mala fide," October 2012.

The World Bank, "A Decade of Action in Transport: An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005," Technical Report, The World Bank, Washington, DC 2007.

Uppal, Yogesh, "The disadvantaged incumbents: estimating incumbency effects in Indian state legislatures," *Public Choice*, July 2009, *138* (1-2), 9–27.

Van der Ploeg, Frederick, "Natural resources: curse or blessing?," *Journal of Economic Literature*, 2011, *49* (2), 366–420.

Wyplosz, Charles, "Fiscal Rules: Theoretical Issues and Historical Experiences," 2012.

Appendix A

Appendix to Chapter 1

A.1 A model of political behavior

There are two parties, A and B, with respective policies on a one-dimensional continuum, X_A and X_B. Without loss of generality, let A by the majority party. The majority party allocates a fixed amount of government resources across K constituencies, assigning γ_k to constituency k, subject to the budget constraint $\sum_{k=1}^{K} \gamma_k = 1$.

Each constituency has two politicians, characterized by an inherent ability $\theta_{i,k}$ where $\theta \in [0,1]$, $\mathbb{E}(\theta) = 0.5$ and $i \in \{A, B\}$. This represents the politician's ability to bring useful government inputs to his constituency. After allocations have been decided by the central party, the value of government inputs received by voters in constituency k is equal to $\gamma_k \cdot \theta_{I,k}$, where I represents the incumbent politician in constituency k. A low ability candidate dissipates the value of government inputs; this could be because he allows them to be stolen, or because he obtains inputs that are not useful to his constituency. All candidates are committed to the policy position of their party.

Voter j in constituency k is characterized by a policy position $X_{j,k}$. Voter j's value of voting for candidate i is determined by a convex function of the cost of a candidate's distance from the voter's optimal policy, and a linear preference for candidate quality, as given by the following utility function:

$$U_{j,k,i} = (X_i - X_{j,k})^2 + \hat{\theta}_{i,k},$$

where $\hat{\theta}_{i,k}$ is a voter's perception of the ability of candidate i in constituency k.

Taking a probabilistic voting approach, the probability that candidate A is elected is given by:

$$P(\text{A wins}) = \Phi\left(-(X_A - X_{M,k})^2 + (X_B - X_{M,k})^2 + \hat{\theta}_{A,k} - \hat{\theta}_{B,k}\right),$$

where $\Phi()$ is the normal c.d.f. and $X_{M,k}$ is the optimal policy of the median voter in constituency k. Candidate ability affects success only if the median voter does not have strong preferences for either party position.

Voters cannot observe a candidate's θ; they can only see $\gamma \cdot \theta$, which is the final value of government inputs received. Voters discount their observation of government inputs received by their prediction of γ_k as follows:

$$\hat{\theta}_{I,k} = \frac{\gamma_k \cdot \theta_{I,k}}{\hat{\gamma}_k}.$$

The party seeks to maximize the probability of re-election, paying a convex cost of deviating from equal provision of inputs to all constituencies.[1] The party's optimization problem is as follows (assuming A controls the government):

$$\max_{\{\gamma_1,\gamma_2,\ldots,\gamma_K\}} \sum_{k=1}^{K} \gamma_k^{\alpha} + P(\text{A wins}|\gamma_k, X_{M,k}, \hat{\theta}_{A,k}, \hat{\theta}_{B,k}),$$

where

$$P(\text{A wins}|\gamma_k, X_{M,k}, \hat{\theta}_{A,k}, \hat{\theta}_{B,k}) = \Phi\left((X_B - X_{M,k})^2 - (X_A - X_{M,k})^2 + \hat{\theta}_{A,k}(\gamma_k) - \hat{\theta}_{B,k}(\gamma_k)\right).$$

Voters estimate candidate ability as:

$$\hat{\theta}_{i,k}(\gamma_k) = \begin{cases} \frac{\theta_{i,k}\cdot\gamma_k}{\hat{\gamma}_k} & \text{if } i \text{ is the incumbent} \\ \mathbb{E}(\theta) & \text{if } i \text{ is not the incumbent} \end{cases}.$$

Denote the median voter's preference of policy A over policy B as:

$$\eta_k = (X_B - X_{M,k})^2 - (X_A - X_{M,k})^2$$

The first order condition defines the relationship between the supply of government inputs across two constituencies:

$$\alpha\gamma_k^{*\alpha-1} + \phi\left(\eta_k + \hat{\theta}_{A,k} - \hat{\theta}_{B,k}\right)\left(\frac{\partial\hat{\theta}_{A,k}}{\partial\gamma_k} - \frac{\partial\hat{\theta}_{B,k}}{\partial\gamma_k}\right) = \alpha\gamma_l^{*\alpha-1} + \phi\left(\eta_l + \hat{\theta}_{A,l} - \hat{\theta}_{B,l}\right)\left(\frac{\partial\hat{\theta}_{A,l}}{\partial\gamma_l} - \frac{\partial\hat{\theta}_{B,l}}{\partial\gamma_l}\right).$$

[1]This cost could reflect a preference for citizen welfare, a political cost of appearing to engage in patronage, or simply an administrative cost of distorting the allocation of inputs from a default level.

The first term indicates the cost of deviating from equal provision. The density function ϕ indicates the marginal electoral return from getting more votes in constituency k: if $|\eta_k|$ is large, then $\phi = 0$ and the party cannot affect the outcome in this location. Provision of inputs will be equal to all non-swing constituencies.

The final term indicates the party's ability to shift voters' perceptions of the quality difference between the candidates. This depends on incumbency, as government spending does not affect perceptions of the non-incumbent candidate:

$$\left(\frac{\partial \hat{\theta}_{A,k}}{\partial \gamma_k} - \frac{\partial \hat{\theta}_{B,k}}{\partial \gamma_k}\right) = \begin{cases} \frac{\theta_{A,k}}{\hat{\gamma}_k} & \text{if A is the incumbent} \\ -\frac{\theta_{B,k}}{\hat{\gamma}_k} & \text{if B is the incumbent} \end{cases}$$

Comparing two aligned constituencies, we get the expression:

$$\alpha \gamma_k^{*\alpha-1} + \phi_k(\cdot)\left(\frac{\theta_{A,k}}{\hat{\gamma}_k}\right) = \alpha \gamma_l^{*\alpha-1} + \phi_k(\cdot)\left(\frac{\theta_{A,l}}{\hat{\gamma}_l}\right),$$

indicating that the candidate in a closer election (indicated by a larger value of $\phi_k(\cdot)$) will receive more resources. Conversely, comparing two non-aligned constituencies, only a sign changes, and the candidate in a closer election will receive *fewer* resources.[2]

Comparing an aligned and a non-aligned constituency, we find:

$$\alpha \gamma_k^{*\alpha-1} + \phi(\cdot)\left(\frac{\theta_{A,k}}{\hat{\gamma}_k}\right) = \alpha \gamma_l^{*\alpha-1} + \phi(\cdot)\left(-\frac{\theta_{B,l}}{\hat{\gamma}_l}\right).$$

The aligned constituency will receive more spending than the non-aligned constituency, but only if one of the two elections is close. The differential is highest if elections are close in both constituencies.

Note that this model does not rely on voters' misunderstanding of party strategy. The party's optimal strategy depends on the distribution of $\hat{\gamma}_k$, but the signs of the comparative statics above are unchanged if we allow fully rational voters, such that $\hat{\gamma}_k = \gamma_k^*$. Voters expect electoral strategy to affect the distribution of government resources (consistent with our results on stock prices), and so they discount the signal received in swing constituencies. But this discounting does not obviate the need for strategic spending - if the party ignores strategy while voters expect it, then swing voters' perceptions of aligned candidates will be biased downward. This result is analogous to the finding that firms manipulate their earnings reports in equilibrium even if investors are aware that

[2] Another implication of the model which we do not exploit is that if the party can observe ability, then higher ability aligned candidates will receive more resources, as the higher θ makes those resources more visible to voters. Conversely, high ability non-aligned candidates receive fewer resources for the same reason.

manipulation is taking place.

A.2 Coalitions and multiple candidates

This section describes how we extend the 2-party empirical strategy in section 1.4 to a situation with more than two parties and coalitions which may change after election results are revealed.

Assume that candidates from N parties contest the election in a given constituency, one of whom is aligned with the ruling party. *margin* is now defined as the scaled vote distance from the aligned candidate to the non-aligned candidate with the highest number of votes:

$$margin_{cst} = \frac{v^{a}_{cst} - v^{maxn}_{cst}}{v^{tot}_{cst}},$$

where v^{maxn}_{cst} is the vote share of the non-aligned candidate with the highest number of votes. *margin* can now be interpreted as the share of votes that would need to be changed to turn an aligned constituency into a non-aligned constituency, or vice versa. As before, *margin* is positive for aligned constituencies, negative for non-aligned constituencies, and elections are closest when $|margin|$ is small.

The formation of coalitions presents a potential source of bias to our identification strategy. Coalitions may be formed before or after an election. If a coalition forms after an election, it is possible that unobserved characteristics of a successful candidate may affect both her likelihood of joining the governing coalition, and economic outcomes in her constituency. For example, if small parties with high ability candidates are more likely to join governing coalitions, Equation 1.1 could overestimate the effect of alignment.

To eliminate this bias, we define coalitions of parties strictly on the basis of information that was known before an election takes place. In many cases, alliances of parties are announced in advance; when possible, we define coalitions according to this information, which we collected from news reports. We then treat coalitions of parties as a single party. When we are unable to find information on coalitions before the election, we predict party alliances on the basis of the previous election in the same state.

In cases where coalitions have shifted during the electoral cycle, this method may incorrectly label coalition parties as non-coalition and vice versa. This contaminates the RD design, biasing our estimates toward zero. The bias is most likely small: we accurately predict candidate alignment in 88% of cases.

A.3 Additional figures and tables

Table A.1: *Placebo regression discontinuity estimates at sample quartiles*

	1	2	3	4	5	6
Aligned (RD)	-0.005	-0.005	0.003	0.001	-0.001	0.001
	(0.006)	(0.004)	(0.003)	(0.008)	(0.006)	(0.004)
Margin	0.166	0.160	-0.031	-0.388	-0.310	-0.077
	(0.219)	(0.202)	(0.035)	(0.279)	(0.216)	(0.048)
Margin * Aligned	-0.120	-0.098	-0.048	0.726	0.631	0.114
	(0.459)	(0.369)	(0.073)	(0.388)*	(0.308)*	(0.043)**
Baseline		-0.014	-0.018		-0.019	-0.018
		(0.004)***	(0.003)***		(0.005)***	(0.003)***
Constant	0.017	0.113	0.199	0.021	0.191	0.199
	(0.006)***	(0.038)***	(0.025)***	(0.008)**	(0.046)***	(0.027)***
N	512	512	3625	617	617	3625
r2	0.17	0.30	0.19	0.14	0.25	0.20

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$. The table shows placebo kernel regression discontinuity estimates of Equation 1.2. Columns 1-3 estimate a discontinuity at the median margin below zero, and columns 4-6 estimate a discontinuity at the median margin above zero. Columns 1 and 4 estimate local linear regressions with state and year fixed effects. Columns 2 and 5 add lagged constituency controls. Columns 3 and 6 estimate polynomial functions across the full range of the margin distribution. Standard errors are clustered at the state-election level.

Table A.2: *RD Estimates of effect of majority alignment on employment growth, polynomial fits of local linear regression*

	1	2	3	4	5	6
Aligned (RD)	0.014	0.015	0.015	0.014	0.017	0.015
	(0.008)*	(0.007)**	(0.007)**	(0.010)	(0.009)*	(0.009)
Margin	-0.573	-0.686	-0.667	-2.124	-1.881	-1.587
	(0.646)	(0.603)	(0.601)	(1.477)	(1.393)	(1.389)
Margin * Aligned	0.942	0.878	0.892	3.831	2.701	2.532
	(0.924)	(0.863)	(0.859)	(2.149)*	(2.022)	(2.012)
Margin^2	-8.913	-12.205	-11.118	-109.892	-89.739	-70.661
	(16.845)	(15.711)	(15.656)	(88.142)	(82.981)	(82.659)
Margin^2 * Aligned	-1.304	4.609	2.728	18.924	44.063	18.657
	(22.830)	(21.303)	(21.198)	(122.251)	(113.924)	(113.122)
Baseline		-0.023	-0.023		-0.023	-0.023
		(0.003)***	(0.003)***		(0.003)***	(0.003)***
Margin^3				-1678.439	-1285.170	-986.221
				(1438.427)	(1351.254)	(1345.630)
Margin^3 * Aligned				2957.299	1889.683	1678.102
				(1973.944)	(1857.160)	(1846.876)
Constant	0.003	0.206	0.200	-0.011	0.195	0.192
	(0.087)	(0.084)**	(0.087)**	(0.087)	(0.085)**	(0.088)**
Weighted	No	No	Yes	No	No	Yes
Controls	No	Yes	Yes	No	Yes	Yes
N	663	663	663	663	663	663
r2	0.13	0.26	0.25	0.13	0.26	0.25

$^*p < 0.10,^{**}p < 0.05,^{***}p < 0.01$. The table shows kernel regression discontinuity estimates of the effect of politician alignment with the state-level governing coalition on annualized constituency log employment growth from 1990-98 and 1998-2005. Polynomial functions of margin are included as controls instead of linear functions. Columns 1-3 present local linear estimates with quadratic controls in margin of victory. Columns 4-6 present local linear estimates with 3rd degree controls in margin of victory. Column 1 is the baseline regression on year and state fixed effects. Column 2 adds lagged constituency controls, and column 3 weights observations by baseline employment. Columns 4-6 follow the same pattern. Standard errors are clustered at the state-election level.

Table A.3: *Event study estimates of CAR in month following election of aligned candidate, full sample rectangular kernel*

	Event study				Placebo Test	
	(1)	(2)	(3)	(4)	(5)	(6)
Aligned	0.024	0.032	0.037	0.047	0.003	0.001
	(0.017)	(0.020)	(0.021)*	(0.025)*	(0.020)	(0.012)
Constant	0.017	-0.048	-0.122	-0.179	-0.018	-0.008
	(0.021)	(0.022)**	(0.050)**	(0.033)***	(0.016)	(0.016)
Fixed Effects	None	State	State,Year	State * Year	None	State * Year
N	710	710	710	710	710	710
r2	0.00	0.08	0.17	0.23	0.00	0.17

$^{*}p < 0.10,^{**} p < 0.05,^{***} p < 0.01$.

The table shows estimates of cumulative abnormal returns of publicly traded firms in the month following an election. The independent variable *aligned* indicates that the winner of the constituency where the firm's headquarters are located is a member of the state-level governing coalition. Returns are measured against a market model with a value weighted index of Indian securities representing the market. The sample consists of all firm-election pairs in constituency- or smaller-sized towns in 1990-2005. Column 1 is the baseline model without fixed effects. Column 2 adds state fixed effects. Column 3 adds state and year fixed effects, and column 4 adds state-year fixed effects. Columns 5 and 6 conduct a placebo test, using the cumulative abnormal returns of firms in the month before the election as the dependent variable.